



universität  
wien

# Diplomarbeit

Titel der Diplomarbeit

Eine Non-Responder-Analyse zur Erfassung des Einflusses  
der Freiwilligkeit auf Testeichung und -validierung

Verfasserin

Elisabeth Unterhofer

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im April 2009

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuer: Univ.-Prof. Dr. Klaus. D. Kubinger

## **Vorwort**

Ich möchte mich anfangs ganz herzlich bei meinem Diplomarbeitsbetreuer Herrn Prof. Dr. Kubinger bedanken, dafür, dass er immer ein offenes Ohr für meine Anliegen und Probleme hatte, die sich im Laufe der Diplomarbeit ergaben und ich mich gut betreut fühlte.

Mein weiterer Dank gilt dem Dekan der Fakultät für Mathematik, Herrn Prof. Harald Rindler, für die Erlaubnis zu dieser Studie, aber auch für die Ermöglichung der Kooperation mit Herrn Prof. Peter Raith und Herrn Dr. Johannes Ploss, die sich als Ansprechpersonen für mich zur Verfügung stellten und somit maßgeblich zum Gelingen dieser Studie beitrugen. Danke auch an die Firma Schuhfried GmbH für die unkomplizierte und verlässliche Zusammenarbeit.

Vielen Dank auch an all jene StudienkollegInnen, die mir mit Rat und Tat zur Seite standen, als Testleiter fungierten oder bei der Datenübertragung behilflich waren. Durch ihre Hilfe fühlte ich mich in meinem Tun unterstützt und wurde auch immer wieder motiviert.

Ein ganz großer Dank geht an sämtliche Freunde und Verwandte, die hinter mir standen und mich unterstützten, besonders hier zu erwähnen: Olaf Tasser und Janina Ziegler, die insbesondere zum Gelingen dieser Diplomarbeit beitrugen. Namentlich erwähnen möchte ich auch jene Freunde, die mich schon lange vor meiner Studienzeit begleitet haben und dies bis zum heutigen Tage tun. Jeder dieser Menschen ist auf eine besondere Art und Weise immer für mich da gewesen und ich wusste, dass ich mich auf sie verlassen kann: Christina Gröber, Monika Wolfsgruber, Michael Ausserhofer, Corinna Tasser, Gertrud Purdeller und Ambach Stephanie. Danke.

Mein innigster Dank geht an meine Familie: Meine Eltern bestärkten mich von Anfang an in meinem Wunsch Psychologin zu werden, durch ihre Unterstützung wurde mir das Studium überhaupt erst ermöglicht; sie und meine Geschwister standen immer hinter mir und glaubten stets an mich. Vielen Dank für alles!

## **Abstract**

For the calibration of most of the psychological diagnostic methods, statistical data from volunteers are taken as the base. This approach implicates some problems, because not later than Karner (2002) it is known, that the dimensionality of this method (validity of the Rasch-model) is also influenced by the personal relevance of each participant's expected consequences connected with the test results. Unfortunately it happens oftentimes, that only self-nominated persons are available for the calibration. In order to assess sample bias because of voluntariness, three responder-subgroups were compared. Therefore a sample of mathematics students was asked to participate three times. The grade of voluntariness was successively affected by personalizing each appeal, which led to more commitment. 30 people were forming the sample ( $n=30$ ). After deciding to participate, all of these 30 people were assigned to one of the three respondergroups. The results of the test-battery regarding all of the three groups were compared. When applying a multivariate variance analysis no significant differences between the three groups could be found. Therefore it is assumed that the voluntariness has no influence on the calibration.

Für die Eichung der meisten psychologisch-diagnostischen Verfahren werden Daten von Freiwilligen herangezogen. Dieses Vorgehen dürfte nicht ganz unproblematisch sein, denn spätestens seit Karner (2002) ist bekannt, dass die Dimensionalität eines Verfahrens (Gültigkeit des Rasch-Models) auch davon beeinflusst wird, ob mit den Ergebnissen für die Teilnehmer persönlich relevante Konsequenzen verbunden sind oder nicht. Leider stehen für die Eichung oft nur selbstnominierte Personen zur Verfügung. Um abschätzen zu können, inwieweit durch Freiwilligkeit eine Verzerrung der Daten zustande kommt, wurden in dieser Studie die Testergebnisse von drei Responder-Untergruppen miteinander verglichen. Dazu wurde eine Stichprobe von Mathematikstudenten dreimal zu einer Testteilnahme aufgefordert. Der Grad der Freiwilligkeit wurde sukzessiv beeinflusst, indem die Aufrufe zunehmend persönlicher gestaltet wurden und sich die Studenten dadurch verstärkt zu einer Teilnahme verpflichtet fühlen sollten. Die Stichprobe bestand aus  $N=30$  Personen, die je nach Zeitpunkt des Entschlusses zur Teilnahme, einer der drei Respondergruppen zugeordnet wurden. Die Ergebnisse dreier Leistungstests dieser Gruppen wurden anschließend miteinander verglichen. Es zeigten sich aber letztlich in den berechneten multivariaten Varianzanalysen keine signifikanten Unterschiede. Somit wird angenommen, dass die Freiwilligkeit keinen Einfluss auf die Testergebnisse und damit auf die Eichung hat.

# Inhaltsverzeichnis

Vorwort .....	2
Abstract .....	3
Tabellenverzeichnis.....	7
Abbildungsverzeichnis .....	8
I. Einleitung.....	9
II. Theoretischer Teil.....	12
1. Definition Non-Responder-Analysen.....	13
2. Probleme aufgrund von Nonresponse .....	13
3. Zunahme und Gründe von Nonresponse.....	14
4. Umgang mit Nonresponse.....	15
4. 1. Analyse von Nonresponse .....	15
4. 2. Möglichkeiten zur Steigerung der Response-Raten .....	16
5. Persönlichkeitseigenschaften und Motive von Respondern und Non-Respondern .....	17
6. Aktive und passive Non-Responder .....	19
III. Empirischer Teil.....	21
7. Zielsetzung und Fragestellung der Untersuchung.....	22
8. Methode.....	23
8. 1. Untersuchungsdesign .....	23
8. 2. Erhebungsinstrumente.....	23
8. 2. 1. Rechnen in Symbolen.....	23
8. 2. 2. Formlogik/Induktives Denken.....	24
8. 2. 3. Family Relation Reasoning Test .....	26
8. 3. Stichprobenkonstruktion .....	27
8. 4. Untersuchungsplanung .....	27
8. 5. Probleme bei der Datenerhebung und Durchführung der Studie .....	30
8. 6. Stichprobe.....	32
8. 6. 1. Beschreibung der Stichprobe I: Gesamtstichprobe .....	32
8. 6. 2. Beschreibung der Stichprobe II - Gesamtstichprobe Aufrufe .....	34
8. 6. 3. Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1 .....	35
8. 6. 4. Beschreibung der Stichprobe II - Männer - Variante 1 .....	36
8. 6. 5. Beschreibung der Stichprobe II - Frauen - Variante 1 .....	36

8. 6. 6. Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1a .....	37
8. 6. 7. Beschreibung Stichprobe II - Geschlecht - Variante 1a.....	37
8. 6. 8. Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1b.....	37
8. 6. 9. Beschreibung Stichprobe II - Geschlecht -Variante 1b.....	37
8. 6. 10. Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1c .....	38
8. 6. 11. Beschreibung Stichprobe II - Geschlecht - Variante 1c.....	38
8. 6. 12. Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 2.....	39
8. 6. 13. Beschreibung Stichprobe II - Männer - Variante 2 .....	39
8. 6. 14. Beschreibung Stichprobe II - Frauen - Variante 2 .....	40
8. 6. 15. Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 2a .....	40
8. 6. 16. Beschreibung Stichprobe II - Geschlecht - Variante 2a.....	40
9. Darstellung der Ergebnisse.....	42
9. 1. Umgang mit Ausreißern .....	42
9. 2. Ergebnisse Variante 1a.....	43
9. 3. Ergebnisse Variante 1b.....	44
9. 4. Ergebnisse Variante 1c.....	45
9. 5. Ergebnis Variante 2a .....	46
9. 6. Deskriptive Analyse der Daten .....	47
10. Diskussion und Ausblick.....	49
11. Zusammenfassung.....	52
12. Literatur.....	54
13. Anhang .....	56
13. 1. Anhang A - Materialien.....	56
13. 2. Anhang B - Tabellen und Abbildungen .....	62
Lebenslauf .....	66

## Tabellenverzeichnis

<b>Tabelle 1:</b> Zeitliche Gliederung der Studie.....	28
<b>Tabelle 2:</b> Gesamtstichprobe Geschlecht .....	32
<b>Tabelle 3:</b> Gesamtstichprobe Art des Mathematikstudiums.....	33
<b>Tabelle 4:</b> Kreuztabelle - Aufrufe/Geschlecht - Variante 1a.....	37
<b>Tabelle 5:</b> Kreuztabelle - Aufrufe/Geschlecht - Variante 1b.....	38
<b>Tabelle 6:</b> Kreuztabelle - Aufrufe/Geschlecht - Variante 1c.....	38
<b>Tabelle 7:</b> Stichprobe II - Gesamtstichprobe Aufrufe - Variante 2a .....	40
<b>Tabelle 8:</b> Kreuztabelle Aufrufe/Geschlecht Variante 2a.....	41
<b>Tabelle 9:</b> Ausreißer .....	43
<b>Tabelle 10:</b> Box-M-Test Variante 1a.....	44
<b>Tabelle 11:</b> Multivariate Tests Variante 1a .....	44
<b>Tabelle 12:</b> Box-M-Test Variante 1b.....	44
<b>Tabelle 13:</b> Multivariate Tests Variante 1b.....	45
<b>Tabelle 14:</b> Box-M-Test Variante 1c.....	45
<b>Tabelle 15:</b> Multivariate Tests Variante 1c .....	46
<b>Tabelle 16:</b> Box-M-Test Variante 2a.....	46
<b>Tabelle 17:</b> Multivariate Tests Variante 2a .....	47
<b>Tabelle 18:</b> Deskriptive Analyse „Extremgruppen“ .....	47

# Abbildungsverzeichnis

<b>Abbildung 1:</b> Ausschnitt eines Beispielitems aus dem Leistungstest Rechnen in Symbolen (Schmotzer et al., 1994). .....	24
<b>Abbildung 2:</b> Ausschnitt eines Beispielitem aus dem Leistungstest Formlogik/Induktives Denken (Hagman & Bratfisch, 2003). .....	25
<b>Abbildung 3:</b> Warming up Items aus dem Familiy Relation Reasoning Test (Poinstingl et al., in Vorb.). .....	26
<b>Abbildung 4:</b> Gesamtstichprobe Altersverteilung .....	33
<b>Abbildung 5:</b> Gesamtstichprobe Anzahl der Semester .....	34
<b>Abbildung 6:</b> Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1 .....	36
<b>Abbildung 7:</b> Stichprobe II - Gesamtstichprobe Aufrufe - Variante 2 .....	39



# **I. Einleitung**

<sup>1</sup>Für die Eichung psychologisch-diagnostischer Verfahren werden hauptsächlich die Daten von freiwilligen Personen herangezogen, Ausnahmen wie BIP (Hossiep, Paschen & Mühlhaus, 2003) oder WIT-2 (Kersting, Althoff & Jäger, 2008) sind selten. Testentwickler müssen sich oft damit zufrieden geben, wenn manche Personen überhaupt dazu bereit sind über eine Routinediagnostik hinaus, noch an zusätzlichen Testungen teilzunehmen oder dies aufgrund von Bezahlung tun. Relativ unbeachtet blieb bis jetzt die Frage, ob diese Freiwilligenstichproben bzw. Stichproben von „Selbstnominierten“ nicht ein systematisch abweichendes Verhalten zeigen. Spätestens seit Karner (2002) ist bekannt, dass die Dimensionalität eines Verfahrens davon abhängt, ob für die Personen mit den Testergebnissen persönlich relevante Konsequenzen verbunden sind oder nicht. Trotzdem liegen, bis auf die Diplomarbeit von Leitner (2008), noch keine empirischen Ergebnisse vor, die das Ausmaß der Verzerrung der Eichung aufgrund von Freiwilligenstichproben abschätzen ließen; auch Auswirkungen auf die Validität sind zu befürchten.

Ein möglicher Ansatz, diese Problematik zu untersuchen, stellt eine Non-Responder-Analyse dar. Da echte Non-Responder (Verweigerer) per Definition zu keiner Teilnahme bereit sind, werden Untergruppen von Respondern verglichen, die sich im Grad der Freiwilligkeit bzw. in der „Schnelligkeit“ der Selbstnominierung unterscheiden. Die Null-Hypothesen einer Non-Responder-Analyse besagt also, dass es keine Unterschiede zwischen „testophilen“ und mehr oder weniger „testophoben“ (verweigernden) Personen gibt. Trifft dies zu, kann man die gewonnen Daten für eine Eichung und Validierung verwenden. Kann diese Hypothese jedoch nicht beibehalten werden, so ist z. B. im Zuge einer Bewerberauswahl zu bedenken, ob der Einsatz des fraglichen psychologisch-diagnostischen Verfahrens eventuell verfehlt ist, da sich die Testperson offensichtlich nur zur Erreichung bestimmter Ziele, in diesem Fall einer Einstellung testen lässt. Die Person könnte in der Bewerbungssituation eine höhere Leistungsmotivation oder bessere Konzentrationsleistung aufzeigen, auch könnten Verfälschungsabsichten oder „Prüfungsängste“ auftreten. So gesehen, müsste die Datenerhebung zur Eichung vieler psychologisch-diagnostischer Verfahren genau in dem „Setting“ erfolgen, in dem es später eingesetzt werden soll.

Leitner (2008) führte im Rahmen ihrer Diplomarbeit eine Non-Responder-Analyse durch. Sie untersuchte dazu eine Stichprobe von Jus- und BWL-Studenten im ersten Semester, indem sie an drei aufeinander folgenden Lehrveranstaltungen einen Aufruf zur Teilnahme an einer Testung startete. Je nachdem, wann sich die Studenten zur Teilnahme meldeten, wurden sie

---

<sup>1</sup> Dieser Teil der Einleitung einschließlich bis zum 2. Absatz, wurde anhand der Unterlagen, zum Tagungsbeitrag von K. D. Kubinger und A. Leitner auf der 7. Tagung der ÖGP 2008 in Linz, erstellt.

einer der drei Respondergruppen zugeordnet. Diese drei Gruppen sollten verschiedene Grade von Freiwilligkeit abbilden. Bei der Testung wurden die Studenten ergänzend gefragt, wie oft sie tatsächlich den Aufruf gehört hatten, sodass gegebenenfalls ihre Zuordnung korrigiert werden konnte. Es wurden aber keine signifikanten Unterschiede zwischen den Gruppen in Hinblick auf die Testergebnisse gefunden, folglich musste der Schluss gezogen werden, dass die Freiwilligkeit keinen Einfluss auf die Testergebnisse und somit auf die Eichung hat.

In ihrer Studie verwendete Leitner (2008) zwei Objektive Persönlichkeitstests. Diese Verfahren dürften aber einen sehr hohen Aufforderungscharakter haben und Akzeptanz bewirken, so dass die Frage offen blieb, ob bei „typischen“ psychologisch-diagnostischen Verfahren, wie Leistungstests, dieselben Ergebnisse zu Tage treten würden. Auch hatte sie bei den Aufrufen sehr zu einer Teilnahme motiviert, in dem die Studenten neugierig auf die Testung gemacht wurden und ihnen eine Rückmeldung der Ergebnisse in Aussicht gestellt wurde. Allerdings stellt eine Rückmeldung einen äußeren Anreiz dar, die Teilnahme zur Testung sollte aber freiwillig erfolgen und nicht durch Versprechungen beeinflusst werden. Kritisch zu sehen bleibt auch der Aspekt, ob „gleichartige Aufrufe“ tatsächlich einen unterschiedlichen Grad an Freiwilligkeit abbilden (Leitner, 2008). Vielmehr sollte durch die sukzessive Steigerung des Aufforderungscharakters der Aufrufe ein unterschiedlicher Grad an Freiwilligkeit erzeugt werden.

Wie oben bereits erwähnt, liegen noch kaum empirische Arbeiten vor, die sich speziell mit der Problematik der Verwendung von Freiwilligendaten für die Eichung psychologisch-diagnostischer Verfahren beschäftigten. Daher folgt am Anfang dieser Diplomarbeit eine kurze theoretische Einführung über die generelle Problematik, die sich aufgrund von Nonresponse in den verschiedensten empirischen Untersuchungen ergibt. Es wird eine Definition für Non-Responder-Analysen gegeben, weiters werden mögliche Gründe der Zunahme von Nonresponse und der Umgang damit diskutiert. Der theoretische Teil wird mit einer Darstellung der Unterschiede zwischen Respondern und Non-Respondern abgeschlossen.

Im empirischen Teil wird das Ziel und das Untersuchungsdesign dieser Studie genauer vorgestellt. Ein besonderes Anliegen war es, die oben angeführten Kritikpunkte an der Studie von Leitner (2008) bei der Untersuchungsplanung zu berücksichtigen. Es erfolgt eine genaue Beschreibung der verwendeten psychologisch-diagnostischen Verfahren sowie der Stichprobe. Anschließend werden die statistischen Analysen und die Ergebnisse wiedergegeben. Im letzten Teil werden die anfänglichen Hypothesen in Bezug auf die resultierenden Ergebnisse diskutiert sowie ein Ausblick für zukünftige Studien gegeben.

## **II. Theoretischer Teil**

## **1. Definition Non-Responder-Analysen**

Das Problem der Non-Responder findet man in allen Arten empirischer Untersuchungen. Zum Beispiel bei schriftlichen Befragungen werden häufig die Antworten der spontan teilnehmenden Personen mit jenen verglichen, die erst nach einer oder mehreren Mahnungen bereit sind den Fragebogen auszufüllen. Sollten sich diese beiden Gruppen systematisch bzgl. einer oder mehrerer antwortrelevanter Variablen unterscheiden, so geht man davon aus, dass Unterschiede zwischen Respondern und endgültigen Verweigerern (echte Non-Responder) in noch größerem Ausmaß bestehen. Sollte es keine Unterschiede zwischen den beiden Gruppen geben, so schließt man daraus, dass Non-Responder das Ergebnis nicht systematisch verfälschen würden (Bortz & Döring, 2006).

Auch im Wiener Self-Assessment zur Studierendenwahl befasst man sich mit dieser Problematik. Man nimmt an, dass die Stichproben von Freiwilligen üblicherweise verzerrt, also nicht repräsentativ für die Grundgesamtheit sind. Daher vergleicht man die Ergebnisse von Freiwilligen, die sich sofort zu einer Testung bereit erklärten mit denen, die sich erst nach wiederholtem Angebot dafür entschieden haben. Der Trend des Unterschiedes in den Testergebnissen wird letztendlich auf die Gruppe jener Personen umgelegt, die an keiner Testung bzw. Beratung teilnehmen (Frebort & Kubinger, 2008).

Wie oben bereits erwähnt, betrifft Nonresponse sämtliche Arten empirischer Untersuchungen. Leider gibt es bis auf die Diplomarbeit von Leitner (2008) noch keine empirische Arbeit, die sich speziell mit der Problematik der Verwendung von Daten von Freiwilligen für die Eichung von psychologisch-diagnostischen Verfahren befasst. Daher soll in den nächsten Kapiteln erläutert werden, inwiefern Nonresponse für empirische Studien ein Problem darstellt, wie versucht wird dem entgegenzuwirken und welche Unterschiede zwischen Respondern und Non-Respondern gefunden wurden.

## **2. Probleme aufgrund von Nonresponse**

Laut Burkell (2003) ist das Response-Problem, ähnlich dem des so genannten „blinden Flecks“. Der blinde Fleck ist ein Punkt auf der Retina, der keinen direkten visuellen Input erhält. Das Sehsystem ergänzt diese fehlenden Informationen durch die umliegenden Zonen, um eine möglichst gute Passung zwischen dem was wir sehen und dem was wir aktuell nicht

sehen können zu erreichen. Dieser Prozess schlägt aber fehl, wenn neue und unerwartete visuelle Informationen genau innerhalb der Grenzen des blinden Flecks liegen, in diesem Fall bleiben potentiell wichtige Informationen unbemerkt.

Forscher sind analog dazu ebenfalls mit einem blinden Fleck konfrontiert und zwar bezüglich der fehlenden Daten, die aufgrund der Verweigerung entstehen: sie erhalten nur wenige oder keine Information über Non-Responder, sollten aber dennoch ein komplettes Bild der untersuchten Population liefern. Daher versucht man, die fehlenden Informationen mit den Daten der Responder zu überbrücken, in der Hoffnung, dass sich Responder und Non-Responder ähnlich sind. Allerdings schlägt dieses Vorgehen fehl, wenn die beiden Gruppen sich unterscheiden, dann wird nämlich lediglich die Population der Responder beschrieben (Burkell, 2003).

Aufgrund von Nonresponse ergeben sich nach Burkell (2003) letztlich zwei Probleme bei der Interpretation von Untersuchungsergebnissen: erstens wird die Stichprobengröße reduziert, so dass die Genauigkeit der Ergebnisse vermindert wird und zweitens entsteht, durch das Ausschließen einer nicht zufälligen Teilmenge der Population, ein Fehler in der Stichprobe. Wenn diese Teilmenge sich von der in der Untersuchung involvierten Personen unterscheidet, dann sind die Ergebnisse der Untersuchung verzerrt.

Einige Forscher behaupten, dass es zu keiner Verzerrung durch Nonresponse kommt, wenn keine systematischen Unterschiede zwischen Respondern und Non-Respondern bestehen. Allerdings ist nach Burkell (2003) zu bedenken, dass es nur keine signifikanten Unterschiede in den gemessenen Eigenschaften gibt. Es besteht aber prinzipiell immer die Möglichkeit, dass sich diese beiden Gruppen in Variablen unterscheiden, die nicht bedacht oder erhoben wurden. Nach Burkell (2003) ist es legitim Untersuchungsergebnisse mit hohen Non-Responder-Raten zu interpretieren, wenn argumentiert werden kann, dass die Responder repräsentativ für die Gesamtpopulation sind.

### **3. Zunahme und Gründe von Nonresponse**

Nach Barclay, Todd, Finlay, Grande und Wyatt (2002, S.106) nahm die Response-Rate in Großbritannien über die Jahre immer stärker ab, die meisten veröffentlichten Studien erzielten mittlerweile eine Teilnehmerzahl von 61 %. Auch für die übrigen westeuropäischen Staaten wird seit den 1950er Jahren eine Zunahme von Nonresponse bei Bevölkerungsumfragen festgestellt (Neller, 2005).

Als Gründe für die Zunahme von Teilnahmeverweigerung, geben Groves und Couper (1998) folgende an: allgemeine gesellschaftliche Entwicklungen (z.B. Stress und Zeitmangel, erhöhte Mobilität), steigendes Bedürfnis nach Datenschutz sowie eine gewisse Ermüdung der Bürger durch die stetig steigende Zahl von Studien im Bereich der Marktforschung.

## **4. Umgang mit Nonresponse**

Das Ignorieren von Nonresponse kann nach Burkell (2003) nie gerechtfertigt sein, da man nie genau wissen kann, ob die Untersuchungsergebnisse durch Nonresponse beeinflusst werden. Liegt die Response-Rate unter 75 % sollte deshalb darauf hingewiesen werden, dass die Untersuchungsergebnisse nur begrenzt generalisierbar sind.

Wenn es aufgrund von Nonresponse zu einer Verzerrung der Untersuchungsergebnisse kommt, so ist es möglich, den Einfluss dieser Verfälschung abzuschätzen: es wird errechnet, wie Non-Responder im ungünstigsten Fall oder durchschnittlich geantwortet hätten. Diese Information kann anschließend dazu genutzt werden, die Genauigkeit, mit der die Ergebnisse angegeben werden zu reduzieren. Eine weitere Möglichkeit besteht darin, Subgruppen miteinander zu vergleichen. Sollten sich Unterschiede zwischen diesen Gruppen ergeben, so kann abgeschätzt werden, inwieweit diese Unterschiede durch Nonresponse bedingt sind. Eine tatsächliche Verzerrung der Daten aufgrund von Nonresponse, könnte auch korrigiert werden: dafür werden meist Responder mit einer zweiten Gruppe verglichen, die entweder Non-Responder inkludiert oder diese repräsentiert. Hierzu gibt es verschiedene Möglichkeiten: zum Beispiel können die Responder in Bezug auf bestimmte Variablen (meist demografische Variablen) mit der gesamten Population verglichen werden, damit kann der Frage nachgegangen werden, ob Responder eine zufällige Stichprobe der Population darstellen (weitere Möglichkeiten siehe Kapitel 4. 1.). Werden signifikante Unterschiede zwischen den Gruppen gefunden, so kann man von einem Stichprobenbias ausgehen. Mit Hilfe einer neuen Gewichtung der Daten, kann man anschließend versuchen, die Unterrepräsentanz der Subgruppen zu berücksichtigen (Burkell, 2003).

### **4. 1. Analyse von Nonresponse**

Porter und Whitcomb (2005) führen vier Zugänge an, um Nonresponse zu analysieren:

- 1) Es werden die Personen miteinander verglichen, die zu einem früheren bzw. späteren Zeitpunkt an einer Untersuchung teilnehmen. Man geht dabei davon aus, dass späte Responder den Non-Respondern sehr ähnlich sind.
- 2) Die Respondergruppe wird mit den Personen verglichen, die sich erst bei der zweiten Anfrage zu einer Teilnahme entschließen.
- 3) Die Respondergruppe wird aufgefordert an einer weiteren Untersuchung teilzunehmen. Die Personen, die an beiden Studien teilnehmen, werden anschließend mit den Personen verglichen, die beim zweiten Mal zu keiner Teilnahme mehr bereit waren.
- 4) Sowohl für Responder als auch Non-Responder liegen schon im Vorfeld Daten bzgl. demografischer Variablen u. ä. vor. Diese Daten werden anschließend genutzt, um Non-Responder zu analysieren.

Deming (1960, zitiert nach Burkell, 2003) beschreibt des Weiteren die Möglichkeit, eine Teilgruppe von Non-Respondern zufällig auszuwählen, von denen anschließend eine hundertprozentige Teilnehmerquote erzielt wird. Anschließend können die Daten dieser konvertierten Non-Responder mit den anfänglichen Respondern verglichen werden, um das Ausmaß der Verzerrung der Nonresponse auf alle in der Studie erhobenen Variablen abschätzen zu können.

Leider wird nicht näher darauf eingegangen, mit welchen Mitteln eine hundertprozentige Teilnehmerquote erreicht werden kann, das Vorgehen wird nach Deming (1960, zitiert nach Burkell, 2003) nur als sehr zeitaufwändig und teuer beschrieben.

#### **4. 2. Möglichkeiten zur Steigerung der Response-Raten**

Es gibt verschiedene Methoden, um Nonresponse zu reduzieren. Schon die Art der Untersuchungsdurchführung führt zu unterschiedlichen Response-Raten. Demnach erzielen face-to-face Kontakte die höchsten Teilnehmerzahlen, telefonischer Kontakt führt zu den nächst höheren Response-Raten, gefolgt von der schriftlichen Kontaktierung via Brief und E-Mail. Ebenfalls scheinen Maßnahmen, wie eine Vorankündigung der Untersuchung, die Verwendung von Erinnerungsschreiben, kleine Geldbeträge u. ä. die Teilnehmerzahl zu vergrößern (Burkell, 2003).

Eine weitere Möglichkeit, die Stichprobengröße deutlich zu erhöhen ist, laut Barclay et al. (2002), eine Gruppe von Non-Respondern telefonisch zu kontaktieren und sie zu überzeugen,



wenigstens eine verkürzte Form des ursprünglichen Fragebogens zu bearbeiten. So ist es auch möglich, im Nachhinein die originalen Daten auf Basis der Antworten dieser Subgruppe von Non-Respondern zu korrigieren.

Neller (2005) konzentrierte sich in ihrer Studie vorwiegend auf die Gruppe der Verweigerer. Dabei handelte es sich hier um jene Personen, die bereits eine Änderung ihrer Meinung über die Teilnahme an der Studie, durch einen Interviewer „vor Ort“ abgelehnt hatten. Diese Personen sollten mit Hilfe eines telefonischen Kontakts doch noch von einer Teilnahme überzeugt werden. Die höchsten Teilnehmerquoten zeigten sich für jene Personen, welche anfangs „Kein Interesse“, „Keine Zeit“ und „Grundsätzliche Verweigerung“ als Gründe der Nicht-Teilnahme angegeben hatten. Interessant war nach Neller (2005) vor allem das Potenzial der grundsätzlich „harten“ Verweigerer, die üblicherweise in vergleichbaren Studien nicht mehr erneut aufgesucht worden wären. Ein Hauptanliegen dieser Untersuchung war es zu prüfen, inwieweit es gerechtfertigt ist, von den konvertierten Verweigerern auf die echten Non-Responder zu schließen. Aufgrund erhobener Basisdaten der meisten Non-Responder (soziodemographische Merkmale, höchster Schulabschluss und politisches Interesse) konnte dieser Frage auch nachgegangen werden. Es zeigte sich aber, dass die konvertierten Verweigerer kaum Gemeinsamkeiten mit den echten Non-Respondern aufwiesen, stattdessen ähnelten sie in fast allen Aspekten sehr viel stärker den übrigen Interviewteilnehmern. Somit dürfte es nicht unproblematisch sein, von den konvertierten Verweigerern auf echte Non-Responder zu schließen (Neller, 2005).

## **5. Persönlichkeitseigenschaften und Motive von Respondern und Non-Respondern**

In der Studie von Carlo, Okun, Knight und de Guzman (2005) wurde an 796 Studenten untersucht, inwieweit die Faktoren Verträglichkeit, Extraversion und die prosoziale Wertmotivation sich freiwillig zu melden (prosocial motivation to volunteer), Einfluss auf freiwilliges Verhalten (ehrenamtliche Tätigkeit) ausüben.

Es zeigte sich, dass von den *Big Five*-Persönlichkeitsfaktoren, die Verträglichkeit die höchste Korrelation mit freiwilligem Verhalten aufwies, gefolgt von Extraversion. Verträglichkeit wies außerdem den stärksten Zusammenhang mit prosozialer Wertmotivation auf, der Zusammenhang zwischen Verträglichkeit und Extraversion war hingegen mäßig.

Der generell höchste Korrelationskoeffizient ( $r=0,38$ ) ergab sich zwischen prosozialer Wertmotivation und freiwilligem Verhalten.

In weiteren Analysen zeigten sich Verträglichkeit und das Geschlecht als die stärksten Prädiktoren für soziale Wertmotivation, außerdem übten Verträglichkeit und Extraversion einen indirekten Effekt auf das freiwillige Verhalten über prosoziale Wertmotivation aus. Wenn Personen hohe Scores auf den beiden Dimensionen Extraversion und Verträglichkeit erzielten, war die prosoziale Wertmotivation am stärksten ausgeprägt. Laut Carlo et al. (2005) kann durch die Berücksichtigung von Motiven, die Beziehung zwischen Eigenschaften und Verhalten besser erklärt werden, z. B. zeigte sich, dass prosoziale Wertmotivation 56 % des Effektes von Verträglichkeit auf freiwilliges Verhalten beeinflusst. Um freiwilliges Verhalten vorhersagen zu können, scheinen daher bestimmte Eigenschaften zwar notwendig, aber nicht hinreichend zu sein (Carlo et al., 2005).

Die Autoren Lonnqvist et al. (2007) untersuchten in zwei verschiedenen Studien, inwieweit sich Responder und Non-Responder hinsichtlich der *Big Five*-Persönlichkeitsfaktoren unterscheiden. In der ersten Studie wurden 158 Militärsoffiziere gebeten an einer schriftlichen Untersuchung teilzunehmen. Für diese Offiziere lagen aus früheren Untersuchungen bereits Daten über deren Ausprägung auf den fünf Persönlichkeitsfaktoren vor.

In der zweiten Studie wurden Mitglieder größerer Familien gefragt, ob sie an einer epidemiologischen Erhebung mitmachen würden. Die teilnehmenden Personen sollten ihre Geschwister hinsichtlich der *Big Five*-Persönlichkeitsfaktoren einschätzen, gleichzeitig sollten sie angeben, ob sie selbst dazu bereit wären an zusätzlichen psychologisch-diagnostischen Verfahren und Interviews teilzunehmen. Die Personen, die dazu bereit waren, wiesen in den Bewertungen ihrer Geschwister höhere Werte auf den Skalen Gewissenhaftigkeit, Extraversion und Verträglichkeit auf. Diese Ergebnisse waren konsistent mit den Ergebnissen der ersten Studie, weiters erzielten die teilnehmenden Personen niedrigere Ausprägungen auf dem Faktor Neurotizismus (Lonnqvist et al., 2007).

Nach Lonnqvist et al. (2007) haben diese Ergebnisse eine große Bedeutung für die Eichung von psychologisch-diagnostischen Verfahren. Verwendet man die Ergebnisse von Freiwilligen für die Normierung eines Tests, so könnte diese verzerrt werden. Daher sollte, laut den Autoren, versucht werden, die Repräsentativität der freiwilligen Stichprobe im Hinblick auf die relevante Stichprobe in Bezug auf wichtige Variablen abzuschätzen.

Responder bzw. konvertierte Verweigerer sind im Gegensatz zu Non-Respondern tendenziell weiblich, jünger (unter 30) und weisen ein höheres Bildungsniveau auf (Neller, 2005). Diese Ergebnisse stehen nicht ganz im Einklang mit denen von Burkell (2003), demnach seien Responder eher älter, in Bezug auf die übrigen Variablen waren die Ergebnisse konsistent. In der Studie von Lyness und Brumit Kropf (2007) zeigte sich ebenfalls, dass das Response-Verhalten vom Geschlecht, aber auch von nationalen Unterschieden beeinflusst wird. Die Response-Rate ist sowohl bei Frauen als auch in Ländern mit einer stärkeren geschlechtlichen Gleichberechtigung höher.

Außerdem tendieren Responder dazu sich mehr für das Untersuchungsthema zu interessieren und gehen eher davon aus, dass die Studienergebnisse etwas bewirken können (Burkell, 2003).

Porter und Whitcomb (2005) konzentrierten sich auf die Gründe der Studentenpopulation, die die Bereitschaft zur Teilnahme an einer Studie bedingen. Demnach nehmen eher Studentinnen sowie Personen mit einem hohen sozialen Engagement an Studien teil. Studierende, die eine finanzielle Beihilfe erhalten, hingegen eher nicht.

## **6. Aktive und passive Non-Responder**

Nach Rogelberg et al. (2003) muss zwischen zwei Gruppen von Non-Respondern unterschieden werden: Passive Non-Responder sind diejenigen Personen, bei denen das Ausbleiben einer Teilnahme nicht bewusst und nicht schon im Vorhinein als klare Entscheidung getroffen wurde. Bei passiven Non-Respondern kommt es schließlich aufgrund von äußeren Umständen zu keiner Teilnahme. Aktive Non-Responder hingegen fällen die Entscheidung, an einer Studie nicht teilzunehmen, bewusst. Personen, die absichtlich an einer Studie nicht teilnehmen, können auch mit bestimmten Mitteln (z. B. Erinnerungsschreiben) zu keiner Mitarbeit mehr bewegt werden.

Wenn es gelingt, bereits vor einer Studie bestimmte Daten über die Stichprobe zu sammeln, dann ist eine Unterscheidung in aktive und passive Non-Responder möglich.

In der Untersuchung von Rogelberg et al. (2003) wurde an 405 Studenten während mehrerer Vorlesungen ein mit einem Code versehener Zufriedenheitsfragebogen ausgeteilt, der von allen bearbeitet wurde. Dieser Fragebogen erfasste u. a. die beiden Persönlichkeitsfaktoren Gewissenhaftigkeit und Verträglichkeit, außerdem wurden die Studenten gefragt, ob sie bereit seien an einer weiteren Studie teilzunehmen. Den Studenten wurden in den nachfolgenden

Wochen zwei weitere Fragebögen zugesendet. Als aktive Non-Responder galten nun jene Studenten, die die weiteren Fragebögen nicht bearbeiteten und beim ersten angaben, dass sie zu keiner weiteren Teilnahme bereit seien (15 %). Passive Non-Responder hingegen lehnten die Möglichkeit einer weiteren Teilnahme nicht von vornherein ab, bearbeiteten aber die nachfolgenden Fragebögen trotzdem nicht mehr.

Die Ergebnisse zeigten, dass aktive Non-Responder im Gegensatz zu den Respondern weniger gewissenhaft zu sein scheinen. Auch zeigten sie die Tendenz die Universität eher verlassen zu wollen und weniger verträglich zu sein. Die generelle Unzufriedenheit der Non-Responder mit der ausführenden Organisation (in diesem Fall der Universität) suggeriert, dass Zufriedenheit einen weiteren möglichen Grund für Nonresponse darstellt. Zwischen passiven Non-Respondern und Respondern wurden hingegen kaum Unterschiede in den untersuchten Variablen gefunden; dies erklären sich die Autoren dadurch, dass die Nicht-Teilnahme von passiven Non-Respondern nur durch äußere Umstände bedingt zu sein scheint (Rogelberg et al., 2003).

### **III. Empirischer Teil**

## 7. Zielsetzung und Fragestellung der Untersuchung

Das Ziel dieser Studie ist es, Untergruppen von Respondern zu vergleichen, die sich in ihrer Bereitschaft zur Testteilnahme bzw. im Grad ihrer Freiwilligkeit unterscheiden. Hierfür werden Mathematikstudenten herangezogen, die dreimal einen Aufruf zur Teilnahme an einer Testung zum logischen Denken erhalten. Die Aufrufe werden immer persönlicher gestaltet, in der Erwartung, dass sich die Studenten dadurch zunehmend zu einer Teilnahme verpflichtet fühlen, so dass tatsächlich der Grad der Freiwilligkeit sukzessiv beeinflusst wird. Die Hauptfragestellung lautet somit:

Kann man mit Hilfe einer Non-Responder-Analyse feststellen, ob bei Normierung und Validierung Freiwilligeneffekte zum Tragen kommen?

H0: Es gibt keinen Unterschied zwischen den Testergebnissen Freiwilliger des ersten Aufrufes und den Ergebnissen der Personen des zweiten und dritten Aufrufes.

H1: Es gibt einen Unterschied zwischen den Testergebnissen Freiwilliger des ersten Aufrufes und den Ergebnissen der Personen des zweiten und dritten Aufrufes.

Des Weiteren werden Unterschiede zwischen den Geschlechtern mittels folgender Nebenfragestellung untersucht:

Gibt es einen Unterschied in den Testergebnissen zwischen den weiblichen und den männlichen Respondergruppen?

H0: Es gibt keinen Unterschied in den Testergebnissen zwischen den weiblichen und den männlichen Respondergruppen.

H1: Es gibt einen Unterschied in den Testergebnissen zwischen den weiblichen und den männlichen Respondergruppen.

Wenn bei der Überprüfung der Hauptfragestellung keine Unterschiede zwischen den Aufrufgruppen festzustellen sind, so spricht nichts explizit dagegen, die an Freiwilligen gewonnenen Daten für die Eichung und Validierung eines psychologisch-diagnostischen Verfahrens zu verwenden. Wenn aber die Hauptfragestellung aufgrund eines signifikanten Unterschiedes angenommen werden kann, so ist die Verwendung des fraglichen

psychologisch-diagnostischen Verfahrens bei Personen, die sich nur zum Erreichen bestimmter Ziele testen lassen, nicht zweckmäßig.

## **8. Methode**

Im Folgenden werden das intendierte Vorgehen, die verwendeten psychologisch-diagnostischen Verfahren sowie die Stichprobe genauer beschrieben.

### **8. 1. Untersuchungsdesign**

In dieser Studie stellen die Ergebnisse der drei verwendeten Leistungstests die abhängigen Variablen dar. Die Testbatterie wird im Kapitel 8. 2. genauer beschrieben.

Als unabhängige Variablen werden die Gruppeneinteilungen zu den verschiedenen Aufrufen sowie das Geschlecht angesehen.

### **8. 2. Erhebungsinstrumente**

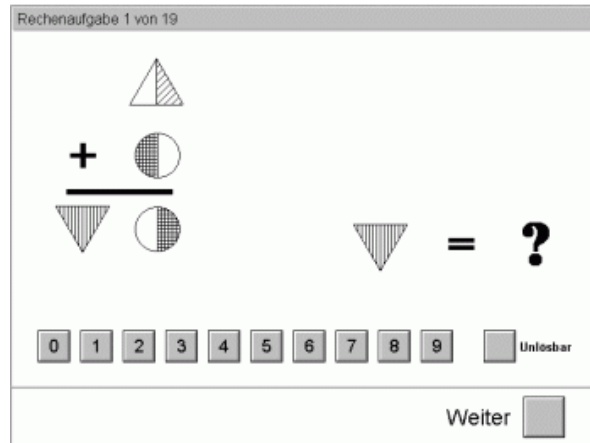
Als Erhebungsinstrumente dienten drei Leistungstests, die alle eine bestimmte Form von Reasoning erheben. „Reasoning ist die Fähigkeit, Gesetzmäßigkeiten oder logisch zwingende Zusammenhänge erkennen und zweckentsprechend verwerten zu können“ (Kubinger, 2006, S. 198). Die drei Verfahren wurden den Testpersonen in der Reihenfolge vorgegeben, in der sie hier präsentiert werden.

#### **8. 2. 1. Rechnen in Symbolen**

Der Leistungstest RIS von Schmotzer, Kubinger und Maryschka (1994) soll die Einhaltung von Restriktionen bei der Bearbeitung von Aufgaben erfassen. Laut Thurstone handelt es sich hierbei, neben induktiven Denken und deduktiven Denken, um den dritten Faktor von Reasoning (Jäger, 1967, zitiert nach Schmotzer et al., 1994).

Die Testpersonen müssen 19 Gleichungen bearbeiten, bei denen die Ziffern durch Symbole ersetzt worden sind, die algebraischen Zeichen für die Rechensymbole werden beibehalten.

Aufgabe ist es, diejenige Ziffer zu finden, die anstatt des gefragten Symbols eingesetzt werden kann, damit man eine mathematisch korrekte Aussage erhält (Schmotzer et al., 1994), siehe auch Abbildung 1.



**Abbildung 1:** Ausschnitt eines Beispielitems aus dem Leistungstest Rechnen in Symbolen (Schmotzer et al., 1994).

Bei RIS handelt es sich um einen Power Test, die Anzahl gelöster Aufgaben dient als Maß für die Fähigkeit der Testperson.

Reliabilität: Die innere Konsistenz ist aufgrund der Geltung des Rasch-Modells gegeben. Die innere Konsistenz nach Cronbach`s-Alpha liegt bei 0,89 (Schmotzer et al., 1994, S. 3).

Gültigkeit: Der Test hat inhaltliche Gültigkeit im Bezug auf schlussfolgerndes Denken nach Thurstone. Die Übereinstimmungsvalidität in Bezug auf den Wiener Matrizen Test (WMT) liegt bei 0,71 (Schmotzer et al., 1994, S. 3).

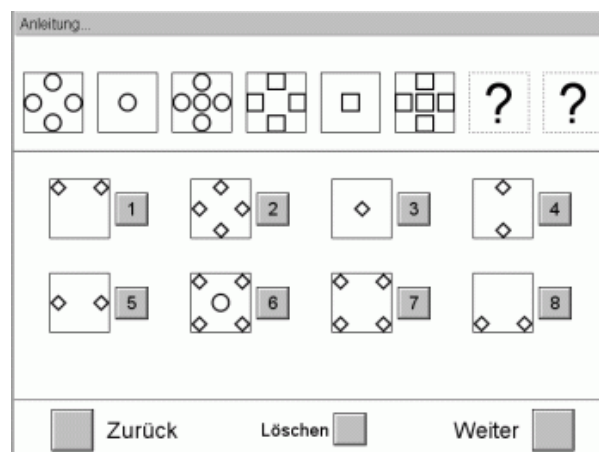
Normen: Die Stichprobe umfasst n=236 Personen im Alter von 16-84. Getrennt für zwei Altersgruppen sind T-Werte und Prozentränge angegeben. Außerdem stehen T-Werte und Prozentränge basierend auf den Daten von 165 PsychologiestudentInnen zur Verfügung (Schmotzer et al., 1994, S. 3).

### **8. 2. 2. Formlogik/Induktives Denken**



FOLO von Hagman und Bratfisch (2003) dient der Erfassung des sprachfreien, logisch-induktiven Denkens. Es handelt sich hierbei um die Fähigkeit, von einem spezifischen Fall eine allgemeine Regel herzuleiten.

Der Test besteht aus 25 Items. Den Testpersonen wird eine Serie von sechs Figuren, die nach einer speziellen Regel aufgebaut ist, dargeboten. Aus acht verschiedenen Möglichkeiten sollen die Testpersonen zwei Figuren auswählen, die die Serie sinnvoll fortsetzen (siehe Abbildung 2) (Hagman & Bratfisch, 2003).



**Abbildung 2:** Ausschnitt eines Beispielitems aus dem Leistungstest Formlogik/Induktives Denken (Hagman & Bratfisch, 2003).

Bei FOLO handelt es sich um einen Speed and Power Test, die Anzahl gelöster Aufgaben dient als Maß für das induktive Denken, wobei eine Aufgabe, dann als gelöst gilt, wenn die Testperson pro Item die zwei korrekten Figuren auswählt.

Reliabilität: Die Split-Half Reliabilität liegt für Personen unterschiedlicher Bildung zwischen 0,89 und 0,91 (Hagman & Bratfisch, 2003, S. 3).

Gültigkeit: „Die logische Gültigkeit ist durch die operationelle Definition des Messinhalts gegeben. Die Begriffsgültigkeit konnte durch Faktorenanalysen nachgewiesen werden. Die prognostische Validität ist für Berufe nachgewiesen, die hohe Anforderungen unter anderem an Logisches Denken stellen“ (Hagman & Bratfisch, 2003, S. 3).

Aus diesen Angaben geht nicht klar hervor, was die Autoren unter der „Begriffsgültigkeit“ verstehen, eventuell handelt es sich um die Konstruktvalidität. Aufgrund dieses Unverständnisses, wurden die Testautoren hier wörtlich zitiert.

Normen: Neben einer schwedischen Stichprobe liegen auch Normen von n=375 österreichischen Erwachsenen vor, diese gibt es aufgegliedert nach Alter und Bildungsgrad (Hagman & Bratfisch, 2003, S. 4).

### 8. 2. 3. *Family Relation Reasoning Test*

Der FRRT (Poinstingl, Kubinger, Skoda & Schechtner, in Vorb.) stellt ein Verfahren zur Erfassung des verbal-schlussfolgernden Denkens dar.

Für den Test als Papier Bleistift Version ist ein freies Antwortformat vorgesehen (Poinstingl et al., in Vorb.), da aber die drei Verfahren in dieser Studie auf PC vorgegeben werden sollten, wurde hier das ursprüngliche Antwortformat, in ein Multiple-Choice Format umgeändert.

Der FRRT ist für ein Alter ab 14 Jahren vorgesehen, da es sich aber bei der Stichprobe dieser Studie ausschließlich um Studenten handelte, wies die vorgegebene Version 25 Items mit höheren Schwierigkeitsparametern auf. Die verwendeten Items waren Rasch-Modell konform (Poinstingl et al., in Vorb.).

Den Testpersonen wird ein kurzer Text präsentiert, der die Beziehung zwischen verschiedenen Familienmitgliedern beschreibt, daraus soll die Verwandtschaftsbeziehung zweier bestimmter Personen abgeleitet werden. Der Testperson stehen dafür sechs Antwortmöglichkeiten zur Verfügung, von der jeweils eine richtig ist (siehe Abbildung 3).

<p>Robin ist der Sohn von Bruno. Bruno ist der Ehemann von Liane.</p> <p>Robin ist _____ von Liane.</p>	<p>1) Sohn 2) Bruder 3) Cousin 4) Großvater 5) Enkelsohn 6) Onkel</p>
<p>Peter ist der Enkelsohn von Marie. Gabi ist die Mutter von Peter. Peter hat die Schwester Anja.</p> <p>Marie ist _____ von Anja.</p>	<p>1) Tochter 2) Mutter 3) Cousine 4) Großmutter 5) Enkeltochter 6) Tante</p>

**Abbildung 3:** Warming up-Items aus dem Familiy Relation Reasoning Test (Poinstingl et al., in Vorb.).

In Abhängigkeit davon, ob bei den Items nach der Verwandtschaftsbeziehung einer Frau oder eines Mannes gefragt wird, werden immer dieselben sechs Antwortmöglichkeiten vorgegeben (siehe Antwortmöglichkeiten Abbildung 3).

Beim FRRT handelt es sich um einen Power Test, die Anzahl gelöster Aufgaben dient als Maß für die Fähigkeit der Testperson (Poinstingl, et al., in Vorb.).

### **8. 3. Stichprobenkonstruktion**

Die Stichprobe sollte sich aus Studenten aus Seminaren bzw. Übungen zusammensetzen. Somit war nämlich die Möglichkeit gegeben, an die Personen sowohl E-Mails zu versenden, als auch sie persönlich aufsuchen zu können. Näheres dazu, ist im Kapitel 8. 4. nachzulesen.

Eine Schwierigkeit bei der Auswahl einer geeigneten Stichprobe war, dass den Studenten, im Gegensatz zur Untersuchung von Leitner (2008), keine Rückmeldung über die Testergebnisse im Aufruf versprochen werden sollte. Eine Rückmeldung stellt nämlich einen äußeren Anreiz dar, die Teilnahme zur Testung sollte aber freiwillig erfolgen und nicht durch Versprechungen beeinflusst werden. Aus zahlreichen empirischen Untersuchungen ist bekannt, dass Personen, die sich für den Untersuchungsgegenstand interessieren, eher freiwillig an Studien teilnehmen als weniger interessierte Personen (Bortz & Döring, 2006). Aus diesem Grund wurde als Stichprobe Mathematikstudenten gewählt, die eine Testbatterie zum logischen Denken bearbeiten sollten, in der Erwartung, dass aufgrund der Fähigkeit im logischen Denken auch ein Interesse an einer Testteilnahme besteht.

### **8. 4. Untersuchungsplanung**

Die Aufrufe zur Testteilnahme sollten sich im Gegensatz zur Studie von Leitner (2008) unterscheiden, indem sie immer persönlicher gestaltet wurden. Der Inhalt an sich sollte dabei aber derselbe bleiben. Für die Umsetzung war vorgesehen, dass die Studenten insgesamt zwei E-Mails im Abstand von zehn Tagen erhalten. Der dritte Aufruf sollte dann in den Seminaren bzw. Übungen selbst stattfinden, wobei bei diesem auch noch zusätzlich die Lehrveranstaltungsleiter die Studenten bitten sollten an der Untersuchung teilzunehmen (E-Mails und persönlicher Aufruf, siehe Anhang A).

Da die gewünschte Stichprobe Mathematikstudenten darstellten, wurde im Mai 2008 um eine Kooperationsbereitschaft der Fakultät für Mathematik für diese empirische Studie angesucht. Folgendes wurde ermöglicht<sup>2</sup>:

- Die Namen und Matrikelnummern der Studenten aus den für die Studie ausgewählten Seminaren wurden an mich weitergeleitet, sodass die ersten beiden Aufrufe an die jeweiligen Unet-Mailadressen gesendet werden konnten.
- Der dritte Aufruf konnte persönlich im Seminar erfolgen, wobei auch die Lehrveranstaltungsleiter ihrerseits die Studenten baten an den Testungen teilzunehmen.
- Es wurde ein Raum für die Testungen an der Fakultät für Mathematik zur Verfügung gestellt, in dem 13 Notebooks des mobilen Testsplatzsystems der Test- und Beratungsstelle des Arbeitsbereiches Psychologische Diagnostik und der Firma Schuhfried GmbH aufgebaut waren. Insgesamt gab es 6 Testtage zu je 6 Stunden Testzeit und 3 Tage zu je 7 Stunden Testzeit.

Die Auswahl der Studenten, die gebeten wurden an der Untersuchung teilzunehmen, erfolgte in Zusammenarbeit mit dem Mathematikprofessor Peter Raith. Es wurde darauf geachtet, Teilnehmer von größeren Parallelveranstaltungen (drittes Semester) und Studenten aus höheren Semestern für die Studie auszuwählen, um möglichst zu verhindern, dass Personen an mehreren der ausgewählten Seminare teilnahmen und somit den persönlichen Aufruf im Seminar zweimal hören würden. Studenten aus dem ersten Semester wurden nicht für die Studie ausgewählt, da aufgrund der Erfahrung von Prof. Raith sehr viele Studenten innerhalb des ersten Monats das Studium wieder aufgeben, somit hätte man den Personen zwar eine E-Mail zukommen lassen können, sie dann aber im Seminar nicht mehr erreicht. Auch gab es Zweifel meinerseits, dass Studienanfänger bereits wissen, dass sie einen universitären Mail-Account besitzen, so dass es wahrscheinlich gewesen wäre, dass viele die E-Mails gar nicht lesen. Die Studie erstreckte sich in einem Zeitraum vom 24.10.08-19.11.08 (siehe Tabelle 1).

**Tabelle 1:** Zeitliche Gliederung der Studie

---

<sup>2</sup> Dank der Unterstützung des Dekans der Fakultät für Mathematik Wien, Herrn Prof. Harald Rindler.

24.10.-2.11.	3.11.-9.11.	10.11.-16.11.	17.11.-19.11.
1. Aufruf via E- Mail	Testungen des 1. Aufrufes (3.11- 5.11)	1. Testungen des 1. Aufrufes (10.11.- 12.11.)	2. Testungen des 3. Aufrufes
	2. Aufruf zur Testteilnahme via E-Mail	3. Aufruf zur Testteilnahme in den Seminaren	

Zu den Testungen des ersten und zweiten Aufrufes konnten sich die Studenten über Internet anmelden. Das Anmeldesystem wurde mit Hilfe der Programmiersprache php geschrieben und als Webseite realisiert, die auf dem universitätseigenem Webpace, der jedem Studierenden zur Verfügung steht, gehostet wurde. Die Daten wurden in der MySql Datenbank der Universität gespeichert. Dieses Anmeldesystem war immer nur für eine bestimmte Zeit abrufbar, beim ersten Aufruf vom 24.10-2.11., beim zweiten Aufruf vom 3.11.-9.11. Diese zeitliche Begrenzung war vorgesehen, um eventuelle Missverständnisse zu vermeiden, z. B. damit eine Person, die sich erst beim zweiten Aufruf dazu entscheidet an der Studie teilzunehmen, sich nicht noch für die Testungen des ersten Aufrufes eintragen kann.

Für die Anmeldung mussten sich die Studenten mit ihrer Unet-Emailadresse und ihrer Matrikelnummer einloggen, somit konnte überprüft werden, welche Person sich zu einer Testung eingetragen hatte und welche nicht. Diese Information wurde jeweils für den nachfolgenden Aufruf benötigt: Es war vorgesehen, dass nur die Personen den zweiten Aufruf erhalten, die sich nicht zur Testung angemeldet hatten. Personen, die sich zwar beim ersten Aufruf anmeldeten, dann aber nicht zur Testung erschienen, erhielten die zweite E-Mail erst später in einer leicht abgeänderten Form (siehe Anhang A). Auch konnte so den Studenten beim dritten Aufruf genau gesagt werden, wie viele Personen aus dem jeweiligen Seminar noch an keiner Testung teilgenommen hatten. Dadurch sollte ebenfalls ein höherer Aufforderungscharakter erzielt werden. Bei der Testung selbst wurden die Teilnehmer mittels eines Codes anonymisiert.

Vor dem persönlichen Aufruf im Seminar wurde den Lehrveranstaltungsleitern der genaue Text bzw. Ablauf zugeschickt (siehe Anhang A). In dieser E-Mail wurden alle nochmals gebeten, die Aufforderung ihrerseits zur Testteilnahme nicht zu vergessen, da erst dadurch Nachdruck bei den Studenten erzeugt werden kann. Auch wurde darauf aufmerksam gemacht, dass sich jeder an den vorgegebenen Inhalt halten sollte, damit es nicht zu einer

unterschiedlichen Beeinflussung kommt. Alle Lehrveranstaltungsleiter sendeten eine Bestätigung über den Erhalt dieser E-Mail.

Die Anmeldung beim dritten Aufruf im Seminar erfolgte über eine Terminliste, auf der sich die Studenten unmittelbar eintragen konnten.

Bei den Testungen selbst wurden die Teilnehmer gebeten keine Information über die Testung an andere weiterzugeben, da dadurch die Ergebnisse nachfolgender Personen eventuell beeinflusst werden könnten; auch sollten so die anderen Studenten durch eventuelle Informationen nicht in ihrer Freiwilligkeit beeinflusst werden (Instruktion siehe Anhang A).

Bei den Testungen des zweiten und dritten Aufrufes wurde darauf geachtet die Personen der richtigen Respondergruppe zuzuweisen, in dem alle gefragt wurden, ob sie die erste E-Mail vor der zweiten gelesen hätten bzw. ob sie die Mails gelesen hätten, bevor der Aufruf im Seminar erfolgte.

## **8. 5. Probleme bei der Datenerhebung und Durchführung der Studie**

Zählte man alle Teilnehmer auf den Listen zusammen, so handelte es sich um 168 Personen. Trotz der Bemühungen keine Überschneidungen zu haben, kamen 13 Personen doppelt vor, sie waren also Teilnehmer in zwei der ausgewählten Seminare. Bei der Versendung der E-Mails wurde daher darauf geachtet, dass diese Personen auch nur jeweils eine E-Mail erhielten. Beim persönlichen Aufruf wurde vermerkt, ob sich diese Personen anmeldeten oder aber sich erst beim zweiten persönlichen Aufruf für eine Teilnahme entschieden.

Bei einem Student war keine Matrikelnummer angegeben, so dass dieser nicht kontaktiert werden konnte, bei einer anderen Person kam eine Fehlermeldung beim Versenden der E-Mail. Da die Matrikelnummer korrekt war, wurde davon ausgegangen, dass diese Person keinen Unet-Account angelegt hatte. Insgesamt erhielten somit 153 Personen die erste E-Mail.

Auf den ersten Aufruf meldeten sich im Zeitraum vom 24.10.-30.10. vier Studenten zu einer Testung an. Dies waren weitaus zu wenig Personen, außerdem konnte nicht abgeschätzt werden, wie viele die E-Mail zwar gelesen hatten, aber sich aufgrund des großen Anmeldezeitraumes erst später anmelden wollten oder vielleicht auch vergaßen sich anzumelden. Daher wurde entschieden, dieselbe E-Mail nochmals, gewissermaßen als Erinnerung zu versenden, jedoch mit folgendem Betreff: „Techn. Problem: erneute

Versendung folgender Mail“. Der Text des ersten Aufrufes wurde wie folgt eingeleitet: „Aufgrund eines technischen Problems beim Versenden der E-Mail am 24.10.08, erhalten alle angeschriebenen Personen nochmals folgende Information:“. Daraufhin meldete sich noch eine Person an. Da die beiden E-Mails denselben Inhalt und damit denselben Aufforderungscharakter hatten und die Studenten im Glauben gelassen wurden, dass nicht alle angeschriebenen Personen die E-Mail erhalten hatten, wurde dies nur als ein Aufruf gehandhabt.

Zwei der fünf angemeldeten Personen erschienen letztlich nicht zur Testung des ersten Aufrufes.

Beim Versenden des zweiten Aufrufes, konnte eine Person nicht mehr angeschrieben werden, es wurde zweimal eine Fehlermeldung angezeigt. Bei einem Student stellte sich heraus, dass die Matrikelnummer falsch war (die Teilnehmerliste war mit Hand geschrieben, daher wurde die Matrikelnummer falsch abgelesen), dieser Person konnte daher auch keine E-Mail mehr gesendet werden. Die zwei Personen, die sich zur ersten Testung angemeldet hatten, aber nicht kamen, erhielten den zweiten Aufruf am Abend ihres eigentlichen Termins, diese meldeten sich aber zu keinem weiteren Testtermin im Verlauf der Studie an. Insgesamt erhielten somit 148 Studenten den zweiten Aufruf.

Vier Personen sendeten eine Begründung via E-Mail, warum eine Testteilnahme nicht möglich sei, damit zeigte sich auch, dass es offenbar gelang den zweiten Aufruf persönlicher zu gestalten bzw. dass der Aufforderungscharakter höher war. Auf den ersten Aufruf kam nämlich keine solche Reaktion.

Beim zweiten Aufruf meldeten sich vier Studenten zu einer Testung an, von denen einer nicht erschien.

Wie im Kapitel 8. 4. beschrieben, wurden vor dem persönlichen Aufruf alle Lehrveranstaltungsleiter nochmals angeschrieben und der genaue Ablauf erläutert. Trotz dieser Bemühungen gestalteten sich die Aufrufe allerdings sehr unterschiedlich:

Bei einem Seminar musste der Aufruf „zwischen Tür und Angel“ erfolgen, so dass ich nicht die volle Aufmerksamkeit der Studenten hatte, außerdem richtete der Lehrveranstaltungsleiter seinerseits keine Bitte an die Studenten zur Testteilnahme.

In einem anderen Seminar hingegen hatte der Lehrveranstaltungsleiter meinen Besuch schon kurz vorher angekündigt, er unterbrach den Unterricht während die Anmeldeleiste herumgereicht wurde, außerdem appellierte er noch zusätzlich mit eigenen Worten an die

Studenten, zur Teilnahme an der Studie. Die Dynamik in diesen beiden Gruppen war völlig unterschiedlich, was man auch an der Zahl der Anmeldungen sah: beim ersten Seminar meldeten sich drei von 20 Teilnehmern an, beim anderen waren es 13 von 17 Personen.

Bei einem weiteren Seminar meldeten sich sechs Personen von 20 an, auch hier appellierte der Lehrveranstaltungsleiter noch zusätzlich an die Studenten.

Am 13.11.08 erfolgten die Aufrufe in den restlichen vier Seminaren, wobei sich hier insgesamt nur noch sechs Personen anmeldeten. Mögliche Gründe für diese geringe Zahl könnten folgende sein: Eines der Seminare wurde in dieser Woche nicht vom eigentlichen Professor, sondern von einer Vertretung abgehalten; eventuell fühlten sich die Studenten bei seiner Bitte um eine Teilnahme weniger dazu verpflichtet. Bei zwei Seminaren, die derselbe Professor abhielt, fanden in dieser Woche keine richtigen Übungen, sondern nur eine Wiederholung statt, da man mit dem Stoff schon sehr weit fortgeschritten war. Zu diesen Seminaren kamen daher viel weniger Studenten als sonst, in einem waren es z. B. nur zwölf von den eigentlichen 30 Teilnehmern, dadurch wurden beim persönlichen Aufruf auch viel weniger Personen erreicht als eigentlich vorgesehen.

Insgesamt meldeten sich beim dritten Aufruf 28 Personen an, vier davon kamen letztlich nicht.

## **8. 6. Stichprobe**

Im Folgenden werden die für die Auswertung herangezogenen Stichproben dargestellt. Zuerst erfolgt eine Beschreibung der Gesamtstichprobe, dann wird die Einteilung zu den verschiedenen Aufrufgruppen erläutert.

### ***8. 6. 1. Beschreibung der Stichprobe I: Gesamtstichprobe***

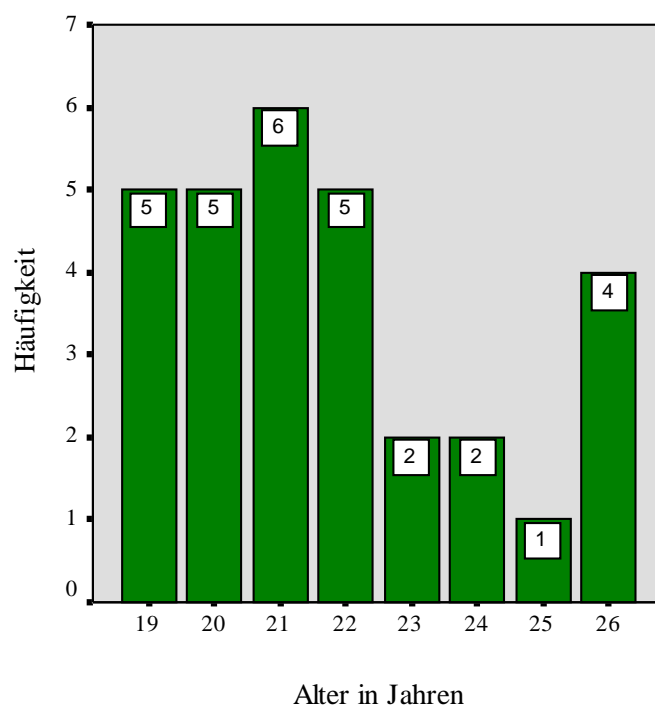
Insgesamt nahmen an der Studie 30 Studenten teil. Davon waren 46,7 % männlich und 53,3 % weiblich (siehe Tabelle 2).

**Tabelle 2:** Gesamtstichprobe Geschlecht



	Häufigkeit	Prozent
<b>männlich</b>	14	46,7
<b>weiblich</b>	16	53,3
<b>Gesamt</b>	30	100

Das Alter der Studenten variierte zum Testzeitpunkt zwischen 19 und 26, der Median lag bei 21 (siehe Abbildung 4).



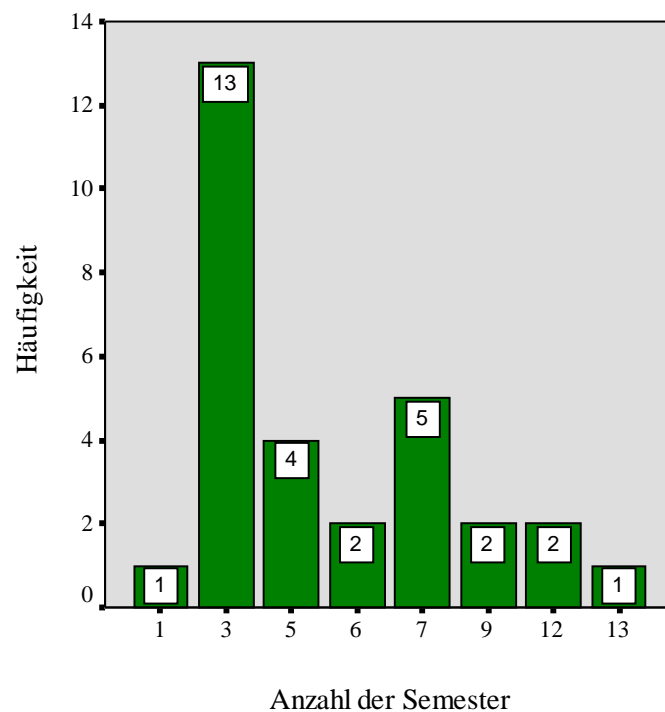
**Abbildung 4:** Gesamtstichprobe Altersverteilung

Von der Gesamtstichprobe studierten fünf Personen Mathematik als Hauptstudium, 24 auf Lehramt und eine Person studierte das Fach sowohl als Hauptstudium als auch auf Lehramt (siehe Tabelle 3).

**Tabelle 3:** Gesamtstichprobe Art des Mathematikstudiums

	Häufigkeit	Prozent
Hauptstudium	5	16,7
Lehramt	24	80,0
Hauptstudium und Lehramt	1	3,3
<i>Gesamt</i>	30	100

Die Semesteranzahl der Studierenden variierte zum Testzeitpunkt zwischen 1 und 13, der Median lag bei 5 Semestern (siehe Abbildung 5).



**Abbildung 5:** Gesamtstichprobe Anzahl der Semester

### 8. 6. 2. Beschreibung der Stichprobe II - Gesamtstichprobe Aufrufe

Die Zuteilung der Teilnehmer zu den einzelnen Respondergruppen erfolgte anhand mehrerer Varianten:

Variante 1) Die Studenten wurden anhand eines Codes zu einer bestimmten Respondergruppe bei der Testung zugewiesen. Dieser Code zeigte an, bei welchem Aufruf sich die Personen angemeldet hatten (z.B. beim Code 0801123 fand die Testung nach dem ersten Aufruf statt,

beim Code 0802133 nach dem zweiten Aufruf usw.). Weiters wurden, wie schon im Kapitel 8. 4. beschrieben, alle Teilnehmer gefragt, ob sie die erste vor der zweiten E-Mail gelesen bzw. die E-Mails vor dem persönlichen Aufruf gesehen haben. Dadurch konnte im Nachhinein die Zuordnung der Personen zu einer anderen Respondergruppe erfolgen (z. B. meldete sich Student A beim persönlichen Aufruf an, gab bei der Testung aber zu verstehen, die beiden E-Mails vorher nicht gesehen zu haben, so wurde die Zuordnung von Respondergruppe 3 in Respondergruppe 1 korrigiert).

Variante 2) Es wurden nur die Studenten in die Stichprobe miteinbezogen, die bei Variante 1 nicht umkodiert werden mussten, also nur die Personen, bei denen der Zeitpunkt der Testung mit der Anzahl der Aufrufe übereinstimmte. Es dürfte nämlich nicht ganz unproblematisch sein, diejenigen Personen, welche sich beim dritten Aufruf anmeldeten und im Nachhinein der 1. Respondergruppe zugeordnet wurden, mit den drei Personen zusammen zu fügen, die sich auf den ersten Aufruf via E-Mail meldeten.

Für die Auswertung ergaben sich aus diesen beiden Varianten folgende vier Teilstichproben:

Variante 1a: Es wurden nur die Teilnehmer des ersten und des dritten Aufrufes in die Berechnung einbezogen.

Variante 1b: Die Teilnehmer des 1. und 2. Aufrufes der Variante 1 wurden zu einer Gruppe zusammengefasst.

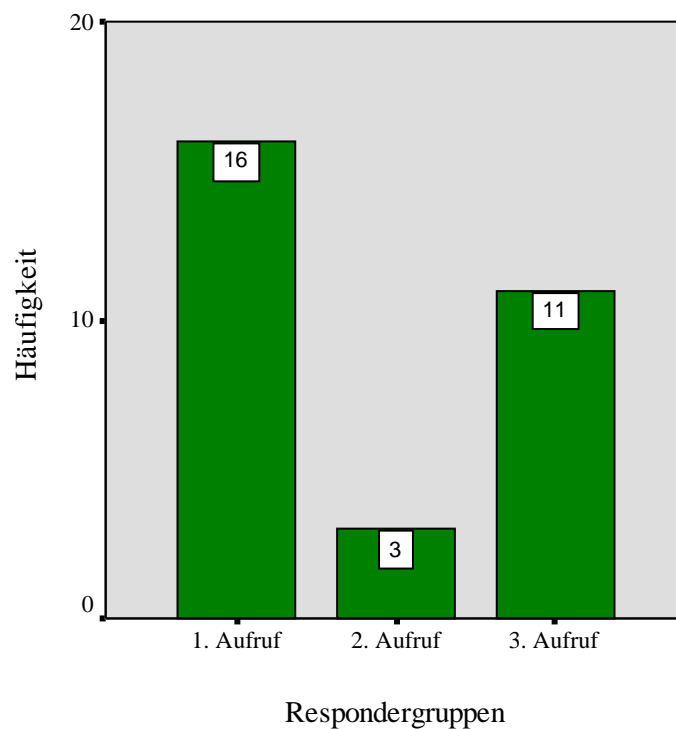
Variante 1c: Die Teilnehmer des 2. und 3. Aufrufes der Variante 1 wurden zu einer Gruppe zusammengefasst.

Variante 2a: Die Teilnehmer des 1. und 2. Aufrufes der Variante 2 wurden zu einer Gruppe zusammengefasst.

### ***8. 6. 3. Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1***

Bei der Variante 1 gaben 53,3 % der Personen an, nur einmal einen Aufruf erhalten bzw. gesehen zu haben, als sie sich zur Teilnahme entschlossen. 10 % entschieden sich nach dem

zweiten Aufruf mitzumachen, 36,7 % nahmen erst nach dem dritten Aufruf an einer Testung teil (siehe Abbildung 6).



**Abbildung 6:** Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1

#### ***8. 6. 4. Beschreibung der Stichprobe II - Männer - Variante 1***

Bei Betrachtung der Teilstichprobe der Männer der Variante 1 ergab sich, dass 64,3 % beim ersten Aufruf an einer Testung teilnahmen, beim zweiten Aufruf waren es 7,1 % und 28,6 % nahmen bei den Testungen des dritten Aufrufes teil (siehe Anhang B).

#### ***8. 6. 5. Beschreibung der Stichprobe II - Frauen - Variante 1***

Bei der Teilstichprobe der Frauen ergab sich nach der Variante 1, dass 43,8 % an einer Testung teilnahmen, nachdem sie den Aufruf einmal gehört hatten, 12,5 % nahmen beim 2. Aufruf teil und 43,8 % beim dritten (siehe Anhang B).

#### **8. 6. 6. Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1a**

Aufgrund der geringen Teilnehmerzahl beim zweiten Aufruf, wurde diese Gruppe in die Berechnung nicht miteinbezogen, somit ergab sich eine Gesamtstichprobengröße von  $n=27$ , wobei 59,3 % dem ersten Aufruf und 40,7 % dem dritten Aufruf zugeordnet wurden (siehe Anhang B).

#### **8. 6. 7. Beschreibung Stichprobe II - Geschlecht - Variante 1a**

Betrachtet man die Gesamtstichprobe der Variante 1a getrennt nach Geschlechtern, so zeigt sich, dass bei den Männern 69,2 % dem 1. Aufruf und 30,8 % dem 3. Aufruf zugeordnet wurden. Die Frauen teilen sich gleichmäßig auf beide Aufrufe auf, mit jeweils 50 %.

Zur Veranschaulichung der Stichprobenaufteilung nach Aufrufen und dem Geschlecht sei folgende Kreuztabelle angeführt (siehe Tabelle 4).

**Tabelle 4:** Kreuztabelle - Aufrufe/Geschlecht - Variante 1a

		Respondergruppe		
		1. Aufruf	3. Aufruf	Gesamt
Geschlecht	männlich	9	4	13
	weiblich	7	7	14
Gesamt		16	11	27

#### **8. 6 . 8. Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1b**

Bei dieser Variante wurden die Personen des ersten und des zweiten Aufrufes zu einer Gruppe zusammengefasst und mit den Teilnehmern des dritten Aufrufes verglichen. Die erste Gruppe umfasst  $n=19$  Studenten, der dritte Aufruf  $n=11$  Personen (siehe Anhang B).

#### **8. 6. 9. Beschreibung Stichprobe II – Geschlecht - Variante 1b**

Betrachtet man die Stichprobe der Variante 1b getrennt nach Geschlechtern, so lässt sich feststellen, dass 71,4 % der Männer dem Aufruf 1 und 2 zugeordnet wurden, 28,6 % dem dritten Aufruf. 56,2 % der Frauen wurden dem Aufruf 1 und 2 zugewiesen, 43,8 % nahmen bei den Testungen des dritten Aufrufes teil. Zur Veranschaulichung siehe nachfolgende Kreuztabelle (siehe Tabelle 5).

**Tabelle 5:** Kreuztabelle - Aufrufe/Geschlecht - Variante 1b

		Respondergruppe		Gesamt
		Variante 1b		
		Aufruf 1+2	3. Aufruf	
Geschlecht	männlich	10	4	14
	weiblich	9	7	16
Gesamt		19	11	30

#### **8. 6. 10. Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1c**

Bei der Stichprobe der Variante 1c wurden die Personen des zweiten und dritten Aufrufes zu einer Gruppe zusammengefasst und mit der 1. Respondergruppe verglichen. 53,3 % der Teilnehmer wurden dem 1. Aufruf zugeordnet, 46,7 % dem Aufruf 2 und 3 (siehe Anhang B).

#### **8. 6. 11. Beschreibung Stichprobe II - Geschlecht - Variante 1c**

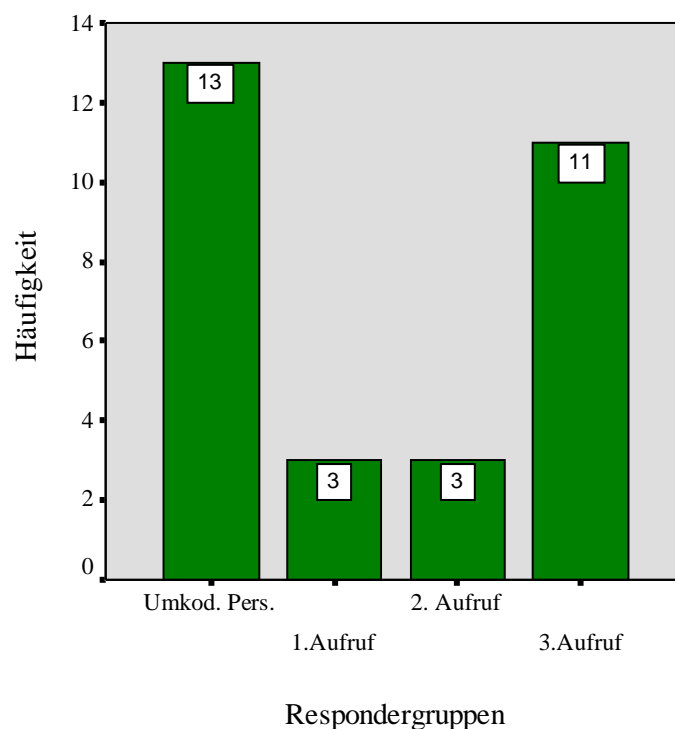
Die Stichprobe der Variante 1c gliedert sich getrennt nach Geschlechtern folgendermaßen: 64,3 % der Männer wurden dem ersten Aufruf zugewiesen, 35,7 % dem Aufruf 2 und 3. Bei den Frauen zählten 43,8 % zum ersten Aufruf und 56,2 % zum Aufruf 2 und 3 (siehe Tabelle 6).

**Tabelle 6:** Kreuztabelle - Aufrufe/Geschlecht - Variante 1c

		Respondergruppe		Gesamt
		Variante 1c		
		Aufruf 1	Aufruf 2+3	
Geschlecht	männlich	9	5	14
	weiblich	7	9	16
Gesamt		16	14	30

#### 8. 6. 12. Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 2

Nach dem ersten Aufruf nahmen  $n=3$  Personen bei einer Testung teil, beim zweiten Aufruf waren es genauso viele. Nach dem persönlichen Aufruf meldeten sich noch  $n=24$  zu einer Testung an, von denen  $n=11$  zur dritten Respondergruppe gehörten und  $n=13$  der ersten Gruppe zugewiesen wurde, also umkodiert wurden (siehe Abbildung 7).



**Abbildung 7:** Stichprobe II - Gesamtstichprobe Aufrufe - Variante 2

#### 8. 6. 13. Beschreibung Stichprobe II - Männer - Variante 2

Betrachtet man die Teilstichprobe der Männer für die Variante 2, so zeigt sich, dass 14,3 % beim ersten Aufruf teilgenommen haben. 7,1 % beim zweiten und 28,6 % entschieden sich beim dritten Aufruf für eine Teilnahme. 50 % der männlichen Population wurden im Nachhinein der 1. Respondergruppe zugeordnet (siehe Anhang B).

#### **8. 6. 14. Beschreibung Stichprobe II - Frauen - Variante 2**

Bei den Frauen teilte sich die Stichprobe bei Variante 2 folgendermaßen: 6,2 % nahmen beim ersten Aufruf teil, 12,5 % beim zweiten. 43,8 % entschlossen sich beim dritten Aufruf für eine Teilnahme, 37,5 % wurden noch der 1. Respondergruppe zugeordnet (siehe Anhang B).

#### **8. 6. 15. Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 2a**

Für die Berechnung wurden die Teilnehmer des 1. und des 2. Aufrufes zusammengefügt, mit der Begründung, dass die E-Mails einen wesentlich geringeren Aufforderungscharakter als der persönliche Aufruf hatten, was man auch an der Anmeldezahl sehen konnte. Die Gruppengröße der beiden Aufrufe einzeln verwendet, wäre außerdem zu gering, um sie mit dem 3. Aufruf vergleichen zu können. Die umkodierte Personen wurden nicht in die Stichprobe einbezogen (siehe Tabelle 7).

**Tabelle 7:** Stichprobe II - Gesamtstichprobe Aufrufe - Variante 2a

	Häufigkeit	Prozent
<b>Aufruf 1+2</b>	6	35,3
<b>Aufruf 3</b>	11	64,7
<b>Gesamt</b>	17	100

#### **8. 6. 16. Beschreibung Stichprobe II – Geschlecht - Variante 2a**

Betrachtet man die Variante 2a getrennt nach Geschlechtern zeigt sich, dass der Aufruf 1 und 2 42,9 % der Männer umfasst, bei Aufruf 3 sind es 57,1 %. Bei den Frauen wurden 30 %



Aufruf 1 und 2 zugeordnet, 70 % nahmen beim dritten Aufruf an einer Testung teil. Zum besseren Verständnis wird auch hier eine Kreuztabelle angeführt (siehe Tabelle 8).

**Tabelle 8:** Kreuztabelle Aufrufe/Geschlecht Variante 2a

		Respondergruppe		Gesamt
		Variante 2a		
		Aufruf 1+2	3. Aufruf	
Geschlecht	männlich	3	4	7
	weiblich	3	7	10
Gesamt		6	11	17

Wie aus dieser Kreuztabelle zu sehen ist, sind die Gruppengrößen, aufgeteilt nach Geschlecht und Aufruf, sehr gering. Daher wurde bei der Berechnung nur die Gesamtstichprobe herangezogen, auf das Geschlecht als zusätzliche unabhängige Variable wurde bei dieser Variante verzichtet.

## 9. Darstellung der Ergebnisse

Alle deskriptiven Auswertungen und Berechnungen wurden mit SPSS Version 11.5 durchgeführt. Für die Varianten 1a-1c (siehe Kapitel 8. 6. 6.- 8. 6. 11.) wurde eine multivariate Varianzanalyse- für die Variante 2a (siehe Kapitel 8. 6. 15.) ein multivariater t-Test gerechnet. Für alle Berechnungen wurde ein Signifikanzniveau von  $\alpha=0,05$  festgelegt.

Um eine Multivariate Varianzanalyse berechnen zu können, müssen im Vorfeld folgende drei Voraussetzungen erfüllt sein:

1. Intervallskalenniveau der abhängigen Variablen
2. Multivariate Normalverteilung der abhängigen Variablen
3. Homogenität der Varianz-Kovarianz-Matrizen der abhängigen Variablen über die Gruppen

Bei den drei verwendeten Leistungstests konnte ein Intervallskalenniveau der Daten angenommen werden. Die Homogenität der Varianz-Kovarianz-Matrix wurde mittels Box-M-Test bei allen vier Varianten überprüft.

### 9. 1. Umgang mit Ausreißern

Bei der Durchsicht der Häufigkeitsverteilungen der Testergebnisse der einzelnen Aufrufe sowie jener der Aufrufgruppen, getrennt nach Geschlecht, zeigte sich, dass das Testergebnis von vier Personen in jeweils einem Verfahren einen Ausreißer darstellte. Als Ausreißer wurde ein Wert definiert, der sich nicht mehr im Intervall von: Mittelwert  $\pm 2 \cdot$  Standardabweichung befand. Zum Beispiel löste Person Nr.8 bei RIS 2 Aufgaben, die untere Grenze des Intervalls liegt bei den Aufrufgruppen aber bei 4,47 und bei den Aufrufgruppen getrennt nach Geschlecht bei 2,206. Somit stellt das Ergebnis dieser Person, einen Ausreißer in diesem Verfahren dar (siehe Tabelle 9).

Häufigkeitsverteilung- Testergebnisse: Aufrufgruppen				Häufigkeitsverteilung- Testergebnisse: Aufrufgruppen getrennt nach Geschlecht		
Pers. Nr.: Anzahl gelöster Aufgaben	Mitt.w.	Std.abw	Mitt.w.- Std.abw. * 2	Mitt.w.	Std.abw.	Mitt.w.- Std.abw. * 2
Pers. Nr.8: RIS: 2	11,31	3,420	4,47	10,33	4,062	2,206
Pers. Nr.16: FRRT: 14	21,38	3,649	14,082	22,00	3,830	14,34
Pers. Nr. 28: FRRT: 10	19,91	4,277	11,356	20,00	4,899	10,202
Pers. Nr. 20: RIS: 6	12,91	3,145	6,62	12,50	4,435	3,63

**Tabelle 9:** Ausreißer

Es stellt sich natürlich die Frage, warum diese Personen im jeweiligen Test ein deutlich schlechteres Ergebnis als der Durchschnitt erzielten. Bei Person Nr. 8 könnte es z. B. sein, dass sie die Instruktion nicht verstanden oder sich etwas anderes von der Testung erwartet hat und sich erst darauf einstellen musste. Das würde auch erklären, warum diese Person nur bei RIS und keinem anderen Verfahren einen Ausreißer darstellte. Genauso könnte man sich mögliche Gründe für die Personen 16 und 20 überlegen, bei allen drei Personen konnte aber kein Hinweis auf eine Unmotiviertheit in Bezug auf die Testung festgestellt werden, sodass man es hätte rechtfertigen können, diese Studenten nicht in die Berechnung mit ein zu beziehen. Anders stellte sich die Situation bei Person 28 dar, diese zeigte wenige Minuten nach Beginn der Testung auf und fragte, ob sie nicht einen Zettel für die Aufgaben verwenden dürfte, sie könne das sonst nicht. Von keiner der anderen getesteten Personen kam eine ähnliche Reaktion. Aufgrund dieser Aussage kamen beide Testleiter zu dem Schluss, dass Person 28 unmotiviert gewesen sein dürfte an der Testung teilzunehmen.

Bei einer weiteren Analyse der Daten (siehe Kapitel 9. 6.) wurde ersichtlich, dass Person 28 auch bei den anderen beiden Verfahren deutlich schlechter als der Durchschnitt der 3. Respondergruppe abschnitt, somit kann bei ihr von einer anhaltenden Unmotiviertheit während der Testung ausgegangen werden. Mit eben dieser Begründung wurde die Person in die nachfolgenden Berechnungen nicht miteinbezogen.

## 9. 2. Ergebnisse Variante 1a

Wie man aus der Tabelle 10 entnehmen kann, ist der Box-M-Test nicht signifikant, somit kann die Homogenität der Varianz-Kovarianz-Matrizen angenommen werden.

<b>Box-M-Test</b>	<b>22,285</b>
<b>F</b>	<b>0,859</b>
<b>df1</b>	<b>18</b>
<b>df2</b>	<b>722,759</b>
<b>Signifikanz</b>	<b>0,630</b>

**Tabelle 10:** Box-M-Test Variante 1a

Die Faktoren AUFRUF und SEX (Geschlecht) und AUFRUF\*SEX (Wechselwirkung zwischen den Faktoren) haben keinen signifikanten Einfluss auf die Ergebnisse aller drei Leistungstests (siehe Tabelle 11).

<b>Effekt</b>		<b>Wert</b>	<b>F</b>	<b>Hypothese</b>	<b>Fehler</b>	<b>Signifikanz</b>
<b>Variante 1a</b>				<b>df</b>	<b>df</b>	
<b>AUFRUF</b>	Hotelling-Spur	0,158	1,053 (a)	3,000	20,000	0,391
<b>SEX</b>	Hotelling-Spur	0,120	0,803 (a)	3,000	20,000	0,507
<b>SEX*AUFUF</b>	Hotelling-Spur	0,111	0,737 (a)	3,000	20,000	0,542

a Exakte Statistik

**Tabelle 11:** Multivariate Tests Variante 1a

Es gibt also bei der Variante 1a keine signifikanten Unterschiede in den Testergebnissen der drei Leistungstests zwischen den Aufrufgruppen und dem Geschlecht, auch nicht bei deren Wechselwirkung.

### 9. 3. Ergebnisse Variante 1b

Wie in Tabelle 12 ersichtlich, ist der Box-M-Test nicht signifikant, somit ist die Homogenität der Kovarianzmatrizen gegeben.

**Tabelle 12:** Box-M-Test Variante 1b

<b>Box-M-Test</b>	24,797
<b>F</b>	0,979
<b>df1</b>	18
<b>df2</b>	684,559
<b>Signifikanz</b>	0,483

Sowohl die beiden Faktoren SEX und AUFRUF, als auch die Wechselwirkung SEX\*AUFUF haben keinen signifikanten Einfluss auf die Testergebnisse in den drei Leistungstests (siehe Tabelle 13).

**Tabelle 13:** Multivariate Tests Variante 1b

Effekt Variante 1b		Wert	F	Hypothese df	Fehler df	Signifikanz
<b>AUFRUF</b>	Hotelling-Spur	0,107	0,819 (a)	3,000	23,000	0,497
<b>SEX</b>	Hotelling-Spur	0,089	0,684 (a)	3,000	23,000	0,571
<b>SEX*AUFUF</b>	Hotelling-Spur	0,117	0,899 (a)	3,000	23,000	0,457

a Exakte Statistik

Es gibt also auch bei der Variante 1b keine signifikanten Unterschiede in den Testergebnissen der drei Leistungstests zwischen den Aufrufgruppen und dem Geschlecht, auch nicht bei deren Wechselwirkung.

#### 9. 4. Ergebnisse Variante 1c

Der Box-M-Test ist nicht signifikant, somit kann auch hier von einer Homogenität der Kovarianzmatrizen ausgegangen werden (siehe Tabelle 14).

**Tabelle 14:** Box-M-Test Variante 1c

<b>Box-M-Test</b>	20,855
<b>F</b>	0,873
<b>df1</b>	18
<b>df2</b>	1323,298
<b>Signifikanz</b>	0,612

Wie in Tabelle 15 zu erkennen ist, haben auch bei der Variante 1c sowohl die Faktoren AUFRUF und SEX als auch deren Wechselwirkung AUFRUF\*SEX keinen signifikanten Einfluss auf die Testergebnisse in den drei Leistungstests.

**Tabelle 15:** Multivariate Tests Variante 1c

Effekt Variante 1c		Wert	F	Hypothese df	Fehler df	Signifikanz
<b>AUFRUF</b>	Hotelling-Spur	0,197	1,507 (a)	3,000	23,000	0,239
<b>SEX</b>	Hotelling-Spur	0,092	0,708 (a)	3,000	23,000	0,557
<b>SEX*AUFRUF</b>	Hotelling-Spur	0,145	1,110 (a)	3,000	23,000	0,365

a Exakte Statistik

Auch bei der Variante 1c ergaben sich also keine signifikanten Unterschiede in den Testergebnissen der drei Leistungstests zwischen den Aufrufgruppen und dem Geschlecht, auch nicht bei deren Wechselwirkung.

## 9. 5. Ergebnis Variante 2a

Wie in Kapitel 8. 6. 16. erwähnt, wurde auf den Einbezug des Geschlechts als 2. Faktor bei dieser Variante verzichtet, da die Gruppengrößen zu gering waren. Da nur noch ein zweifachgestufter Faktor vorliegt, handelt es sich bei dieser Berechnung um einen multivariaten t-Test und nicht um eine Varianzanalyse.

Die Homogenität der Kovarianzmatrizen ist gegeben, da der Box-M-Test auch hier wiederum nicht signifikant ist (siehe Tabelle 16).

**Tabelle 16:** Box-M-Test Variante 2a

<b>Box-M-Test</b>	10,404
<b>F</b>	1,269
<b>df1</b>	6
<b>df2</b>	706,832
<b>Signifikanz</b>	0,269

Der Faktor AUFRUF hat keinen signifikanten Einfluss auf die Testergebnisse in den drei Leistungstests (siehe Tabelle 17).

**Tabelle 17:** Multivariate Tests Variante 2a

Effekt		Wert	F	Hypothese df	Fehler df	Signifikanz
<b>Variante 2a</b>						
<b>AUFRUF</b>	Hotelling-Spur	0,025	0,099(a)	3,000	12,000	0,959

a Exakte Statistik

Auch in der Variante 2a ergaben sich keine signifikanten Unterschiede in den Testergebnissen der drei Leistungstest zwischen den beiden Aufrufgruppen.

## 9. 6. Deskriptive Analyse der Daten

Aufgrund der geringen Stichprobengröße wurde noch eine deskriptive Analyse der Daten vorgenommen, um so evtl. Tendenzen feststellen zu können (siehe Tabelle 18). In diese Analyse wurden nur diejenigen Gruppen miteinbezogen, welche sich wahrscheinlich im Grad der Freiwilligkeit bzw. im Aufforderungscharakter des Aufrufes am deutlichsten unterschieden haben: Zum einen handelt es sich dabei um die 1. Respondergruppe, die nur die Personen umfasst, die sich auf die erste E-Mail gemeldet haben und anschließend an der Testung teilnahmen. Zum anderen wurden im Nachhinein die Personen herausgesucht, die der 3. Respondergruppe angehörten und Teilnehmer des Seminars 2 waren, weiters wurde noch Person Nr. 28 in die Analyse miteinbezogen.

Wie schon im Kapitel 8. 5. erwähnt, gestaltete sich der persönliche Aufruf sehr unterschiedlich. Die hier einbezogene 3. Respondergruppe stammt aus jenem Seminar, in dem der Lehrveranstaltungsleiter am meisten versuchte die Studenten von einer Teilnahme zu überzeugen. Es ist davon auszugehen, dass sich diese 3. Responder im Vergleich zu den Teilnehmern der anderen Seminare wahrscheinlich am stärksten zu einer Teilnahme verpflichtet fühlten.

Die Problematik mit Person 28 wurde bereits im Kapitel 9. 1. diskutiert, im Prinzip stellt sie eigentlich diejenige Art von Personen dar, welche mit dem 3. Aufruf erreicht werden sollten, nämlich die, die eigentlich unmotiviert sind an einer Testung teilzunehmen, aber aus dem Gefühl der Verpflichtung, Gruppenzwang o. ä. es trotzdem tun.

**Tabelle 18:** Deskriptive Analyse „Extremgruppen“

		Anzahl gelöster Aufgaben RIS	Anzahl gel. Aufgaben FOLO	Anzahl gel. Aufgaben FRRT
<b>Resp.gr.1</b> <b>n=3</b>	<b>Mittelwert</b>	14,33	22,33	23,00
	<b>Standardabw.</b>	2,082	0,577	1,732
	<b>Varianz</b>	4,333	0,333	3,000
<b>Resp.gr.3<sup>3</sup></b> <b>Seminar 2</b> <b>n=3</b>	<b>Mittelwert</b>	14,00	21,67	21,67
	<b>Standardabw.</b>	2,000	2,309	2,517
	<b>Varianz</b>	4,000	5,333	6,333
<b>Pers. 28</b>		10	18	10

Betrachtet man die Mittelwerte der einzelnen Gruppen, so zeigt sich, dass die Respondergruppe 1 etwas höhere Mittelwerte in allen 3 Leistungstests aufweist, allerdings ist der Unterschied sehr gering und ist beim FRRT mit nur einer gelösten Aufgabe mehr, noch am deutlichsten. Es zeigt sich also auch beim Vergleich dieser beiden „Extremgruppen“ keine deutliche Tendenz.

Nur bei Person 28 lassen sich wirklich deutliche Unterschiede erkennen, diese löste bei RIS vier Aufgaben weniger als die restliche 3. Respondergruppe, beim FRRT sind es sogar 11. Damit lässt sich nochmals verdeutlichen, dass der Ausschluss dieser Person von der Berechnung gerechtfertigt ist, da sie offensichtlich über die gesamte Testung hinweg unmotiviert gewesen sein dürfte.

<sup>3</sup> Person Nr.28 gehörte auch der 3. Respondergruppe aus dem Seminar 2 an und wurde daher von diesen Berechnungen ausgeschlossen.



## 10. Diskussion und Ausblick

Bei allen vier Varianten konnte kein signifikanter Unterschied in den Testergebnissen der drei Leistungstests zwischen den Aufrufen festgestellt werden. Somit gab es keinen Einfluss der Freiwilligkeit auf die Testergebnisse.

Auch die Nebenfragestellung muss so beantwortet werden, dass bei allen vier Varianten keine signifikanten Unterschiede zwischen den Geschlechtern vorhanden sind und damit auch keine Effekte seitens der Freiwilligkeit auf die Ergebnisse.

Es scheinen also keine Unterschiede in den Leistungen zwischen „testophilen“ und „testophoben“<sup>4</sup> Personen - und damit vermutlich auch Non-Respondern vorhanden zu sein. Die Ergebnisse lassen also den Schluss zu, dass die an Freiwilligen gewonnenen Daten für die Eichung bzw. auch für die Validierung psychologisch diagnostischer Verfahren geeignet sind. Damit stehen diese Ergebnisse im Einklang mit denen von Leitner (2008), die anhand einer deutlich größeren Stichprobe von  $n=90$ , ebenfalls keine signifikanten Unterschiede zwischen den verschiedenen Aufrufgruppen feststellen konnte. Allerdings wurde in der jetzigen Studie der Aufforderungscharakter der Aufrufe sukzessiv verändert, was bei Leitner (2008) nicht der Fall war. Leider war die Stichprobengröße dieser Untersuchung deutlich zu klein, so dass die Aussagekraft der Ergebnisse vermindert wird, eine Generalisierbarkeit ist daher sicher nicht gegeben.

Gründe für die generell geringe Bereitschaft zur Testteilnahme dürften u. a. folgende gewesen sein:

- Sowohl ein Student, als auch der Dekan der Fakultät für Mathematik gaben an, dass viele Mathematikstudenten generell unmotiviert seien, freiwillig zusätzlichen Aufgaben nachzukommen.
- Als weitere Gründe wurden von den Studenten Stress und die Dauer der Testung (max. 75 min) angegeben.
- Bei den Testungen des 3. Aufrufes zeigte sich auch, dass sehr viele Teilnehmer die E-Mails nicht gelesen hatten (13 von 24 Personen). Die Erreichbarkeit der Studenten via Unet-Account dürfte folglich nicht unproblematisch sein.

---

<sup>4</sup> Die Begriffe „testophil“ und „testophob“ entstammen den Unterlagen, zum Tagungsbeitrag von K. D. Kubinger und A. Leitner auf der 7. Tagung der ÖGP 2008 in Linz.

- Auch das Versprechen einer Rückmeldung dürfte einen Einfluss auf die Bereitschaft zur Teilnahme haben. Der Aufforderungscharakter ist aufgrund der Rückmeldung sicher höher, allerdings wird dadurch aber, wie in Kapitel 8. 3. angeführt, die Freiwilligkeit beeinflusst.

Als Kritikpunkt an dieser Studie ist sicherlich zu sehen, dass die Gründe für die Teilnahme bzw. die Nicht-Teilnahme an der Testung nicht erhoben wurden. Es bleibt somit offen, ob die Zuteilung zu den einzelnen Respondergruppen und damit die Unterscheidung zwischen „testophil“ und „testophob“ allein aufgrund des Anmeldezeitpunktes gerechtfertigt ist. Eine mögliche Lösung dieses Problems könnte das Vorgehen von Rogelberg et al. (2003) darstellen, nämlich die Unterscheidung in aktive und passive Non-Responder, vgl. dazu auch Kapitel 6. Die Autoren stellten in ihrer Studie auch nur Unterschiede zwischen aktiven nicht jedoch zwischen passiven Non-Respondern und Respondern fest. Nachfolgende Untersuchungen sollten diesen Punkt daher berücksichtigen, sodass am Ende einer Untersuchung klar ist, dass nur Personen, die gern und freiwillig an einer Studie teilnehmen mit jenen verglichen werden, die sich aktiv und nicht aufgrund äußerer Umstände gegen die Teilnahme entschieden haben. Betrachtet man allerdings den Aufwand, den die Autoren Rogelberg et al. (2003) betrieben haben, um die Unterscheidung in aktive und passive Non-Responder vornehmen zu können, so ist klar, dass eine solche Untersuchung mit Sicherheit zweckdienlich wäre, aber den Rahmen einer Diplomarbeit wahrscheinlich sprengen würde.

Sowohl bei Leitner (2008) als auch in dieser Studie, wurden Studierende als Stichprobe ausgewählt. Im Gegensatz zu Leitner (2008), die Studienanfänger testete, wurde in dieser Studie versucht, aus Gründen, die in Kapitel 8. 4. erläutert wurden, dritt- und höhersemestrige Studenten von einer Teilnahme zu überzeugen. Unüberlegt war bei dieser Untersuchung allerdings, dass vielleicht gerade Studenten (speziell nicht Studienanfänger) eine zu homogen unmotivierte Teilpopulation darstellen und aus diesem Grund für eine solche Studie ungeeignet sind. Einerseits herrscht unter Studenten teilweise die Einstellung vor, mit möglichst geringem Aufwand möglichst viel erreichen zu wollen, sie zeigen also keine hohe Leistungsmotivation; andererseits aber werden insbesondere Studenten teilweise überladen mit Anfragen zur Teilnahme an Befragungen, sodass daher auch mit einer gewissen Ermüdung zu rechnen ist, die sich vielleicht auch in einer generellen Unmotiviertheit äußert, an solchen Studien teilzunehmen. Nachfolgende Untersuchungen sollten daher an anderen

Teilpopulationen vollzogen werden, um feststellen zu können, ob sich hier Freiwilligeneffekte ergeben.

## 11. Zusammenfassung

Spätestens seit Karner (2002) ist bekannt, dass die Gültigkeit des Rasch-Models auch davon abhängig zu sein scheint, ob für die Testpersonen mit den Ergebnissen persönlich relevante Konsequenzen verbunden sind oder nicht. Es ergibt sich daraus die Frage, inwieweit es gerechtfertigt ist, Daten, die an freiwilligen Personen gewonnen wurden, für die Eichung psychologisch-diagnostischer Verfahren zu verwenden. Um diese Frage zu klären, wurde in dieser Studie versucht eine Non-Responder-Analyse durchzuführen, indem Untergruppen von Respondern, die sich im Grad der Freiwilligkeit bzw. Schnelligkeit der Selbstnominierung unterschieden, verglichen wurden: Es erfolgten drei Aufrufe zu einer Teilnahme an einem Leistungstest, die einen zunehmend höheren Aufforderungscharakter hatten. Dadurch sollte untersucht werden, ob Unterschiede zwischen den Testergebnissen Freiwilliger des ersten Aufrufes und den Ergebnissen der Personen des zweiten und dritten Aufrufes vorhanden sind. Außerdem interessierte die Frage, ob sich männliche und weibliche Respondergruppen in ihren Testergebnissen unterscheiden.

Für die Studie wurden Mathematikstudenten als Stichprobe ausgewählt, die zweimal via E-Mail und einmal im Seminar um eine Teilnahme gebeten wurden. Den Studenten sollte im Vorfeld keine Rückmeldung versprochen werden, um die Freiwilligkeit nicht zu beeinflussen, daher wurde versucht durch die Vorgabe von Tests zum logischen Denken das Interesse an einer Teilnahme zu wecken. Die Testbatterie setzte sich aus Rechnen in Symbolen (Schmotzer et al., 1994), Formlogik/Induktives Denken (Hagman & Bratfisch, 2003) und dem Familiy Relation Reasoning Test (Poinstingl et al., in Vorb.) zusammen und wurde am PC vorgegeben.

Insgesamt nahmen von 153 angeschriebenen Studenten 30 an der Studie teil, davon waren 46,7 % männlich und 53,3 % weiblich. Das Alter lag zwischen 19 und 26 Jahren. Nur jeweils drei Personen meldeten sich zu den beiden Aufrufen via E-Mail an, die restlichen 24 entschieden sich erst beim Aufruf im Seminar für eine Teilnahme. Anhand von vier Varianten wurden die Personen den einzelnen Respondergruppen zugewiesen. Für diese Varianten erfolgte mittels multivariater Varianzanalysen bzw. multivariaten t-Test anschließend die Hypothesenprüfung. Es zeigten sich keine signifikanten Unterschiede, sowohl zwischen den Aufrufen, als auch zwischen den Geschlechtern und auch bei der Wechselwirkung beider Faktoren. Somit hatte der unterschiedliche Grad der Freiwilligkeit keinen Einfluss auf die Testergebnisse der drei Leistungstests.

Die Ergebnisse lassen folglich den Schluss zu, dass die an Freiwilligen gewonnenen Daten zur Eichung und Validierung eines psychologisch-diagnostischen Verfahrens verwendet werden können. Leider war die Stichprobengröße dieser Studie sehr gering, so dass die Aussagekraft dadurch deutlich vermindert wird, aber auch Leitner (2008) fand bei ihrer Studie mit einer Stichprobengröße von  $n=90$  keine signifikanten Unterschiede zwischen den Respondergruppen.

Mögliche Gründe für die geringe Bereitschaft zur Testteilnahme könnten Zeitmangel, die Dauer der Testung, die schlechte Erreichbarkeit der Studenten via E-Mail, das Nicht-in-Aussicht-Stellen einer Rückmeldung und auch die nachträglich angeführte Unmotiviertheit von Mathematikstudenten, freiwilligen Aufgaben nachzukommen, sein.

Nachfolgende Untersuchungen sollten auf die Unterscheidung zwischen passiven und aktiven Non-Respondern eingehen (Rogelberg et al., 2003), also die Gründe der Teilnahme bzw. Nicht-Teilnahme an der Testung erheben, um so die Zuweisung zu den Gruppen „testophil“ und „testophob“ rechtfertigen zu können. Ein weiteres Problem könnte sein, dass Studenten vielleicht eine zu einheitlich unmotivierte Teilpopulation, in Bezug auf eine Untersuchungsteilnahme, darstellen. Diese Unmotiviertheit könnte durch die Einstellung mancher Studierenden, aber auch aufgrund einer Ermüdung, durch die Vielzahl an Aufforderungen zur Teilnahme an Befragungen ect. zustande kommen. Zukünftige Studien sollten sich daher auf andere Teilpopulationen konzentrieren.

## 12. Literatur

- Barclay, S., Todd, C., Filay, I., Grande, G. & Wyatt, P., (2002). Not another questionnaire! Maximizing the response rate, predicting non-response and assessing non-response bias in postal-questionnaire studies of GPs. *Family Practice*, 19, 105-111.
- Bortz J.& Döring N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Heidelberg: Springer Medizin.
- Burkell, J. (2003). The dilemma of survey nonresponse. *Library & Information Science Research*, 25, 239-263.
- Carlo G., Okun M. A., Knight G. P. & de Guzman, M. R. T. (2005). The interplay of traits and motives on volunteering: agreeableness, extraversion and prosocial value motivation. *Personality and Individual Differences*, 38, 1293-1305.
- Freibort M.& Kubinger K. D. (2008). Qualitätsansprüche an ein Self-Assessment zur Studienwahlberatung: Der Wiener Ansatz. In H. Schuler & B. Hell (Hrsg.), *Studierendenauswahl und Studienentscheidung* (S.95-101). Göttingen: Hogrefe.
- Groves, R. M. & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: Wiley.
- Hagman, E. & Bratfisch, O. (2003). Formlogik/ Induktives Denken (FOLO) [Software und Manual]. Mödling: Dr. G. Schuhfried GmbH.
- Hossiep, R., Paschen, M., & Mühlhaus, O. (2003). *Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP)* (2. Aufl.). Göttingen: Hogrefe.
- Karner, T. (2002). The volunteer effect of answering personality questionnaires. *Psychologische Beiträge*, 44, 42-49.
- Kersting, M., Althoff, K. & Jäger, A. O. (2008). *Wilde-Intelligenz-Test 2 (WIT-2)* (1. Aufl.). Göttingen: Hogrefe.
- Kubinger, K. D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.
- Leitner, A. (2008). *Der Einfluss der Freiwilligkeit auf die Normierung/Eichtabellen*. Unveröffentlichte Diplomarbeit, Universität Wien.
- Lonnqvist, J. E., Paunonen, S., Verkasalo, M., Leikas, S., Tuulio-Henriksson, A. & Lonnqvist, J. (2007). Personality characteristics of research volunteers. *European Journal of Personality*, 21 (8), 1017-1030.
- Lyness, K. S. & Brumit Kropf, M. (2007). Cultural values and potential nonresponse bias. A multilevel examination of cross-national differences in mail survey response rates. *Organizational Research Methods*, 10, 210-224.
- Neller K. (2005). Kooperation und Verweigerung: Eine Non-Response-Studie. ZUMA-

*Nachrichten*, 57, 9-36.

Poinstingl, H., Kubinger, K., Skoda, S. & Schechtner, C.: *Familiy Relation Reasoning Test*.  
In Vorbereitung.

Porter, S. R. & Whitcomb, M. E. (2005). Non-response in student surveys: The role of demographics, engagement and personality. *Research in Higher Education*, 46, 127-152.

Rogelberg, S. G., Conway, J. M., Sederburg, M. E., Spitzmüller, C., Aziz, S. & Knight, W. E. (2003). Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology*, 88 (6), 1104-1114.

Schmotzer, C., Kubinger, K. D. & Maryschka, C. (1994). Rechnen in Symbolen (RIS)  
[Software und Manual]. Mödling: Dr. G. Schuhfried GmbH.

## **13. Anhang**

### **13.1. Anhang A - Materialien**

- 1. Aufruf zur Testteilnahme via E-Mail
- 2. Aufruf zur Testteilnahme via E-Mail
- Aufruf zur Testteilnahme via E-Mail für Personen, die sich zur 1. Testung angemeldet hatten und nicht erschienen waren
- 3. Aufruf zur Testteilnahme im Seminar
- Instruktion Testung



Betreff: Teilnahme an Testung zum logischen Denken

Liebe Studierende,

an der Fakultät für Mathematik finden im Rahmen einer Diplomarbeit aus Psychologie, PC-Testungen zum logischen Denken statt. Für die Studie werden Mathematikstudenten (sowohl Hauptstudium, als auch LAK) gesucht, die an einer Testung teilnehmen.

Zeit: 3.-5. November, 11-17 Uhr- genaue Termine siehe Internetseite unten

Dauer: max.75 min

Ort: Raum Nr. C210 UZA4

Anmeldung (und evt. Ummeldung): läuft vom 24.10-2.11.

unter: [www.unet.univie.ac.at/~a0307997/](http://www.unet.univie.ac.at/~a0307997/)

Log in: mit Ihrer Unet E-Mail Adresse + Matrikelnummer

Ihre Testergebnisse werden selbstverständlich vertraulich behandelt und nicht an Dritte weitergegeben; für die Teilnahme möchte ich mich schon im Voraus herzlich bedanken!

Mit freundlichen Grüßen

Elisabeth Unterhofer

P.S. Sollten Sie Probleme bei der Anmeldung haben, bitte geben Sie mir Bescheid. Bei erfolgreicher Anmeldung, erhalten Sie eine automatische Bestätigung via E-Mail.

Betreff: Bitte um Teilnahme an Testung zum logischen Denken

Liebe/r Frau/Herr XY

Vor einigen Tagen haben Sie eine E-Mail erhalten, mit dem Hinweis, dass an der Fakultät für Mathematik, Testungen zum logischen Denken im Rahmen einer Diplomarbeit aus Psychologie durchgeführt werden und dafür Mathematikstudenten gesucht werden. Weil Sie sich zur keiner Testung angemeldet haben, möchte ich Sie erneut um eine Teilnahme bitten:

Zeit: 10.-12. November, 11-17 Uhr- genaue Termine siehe Internetseite unten

Dauer: max.75 min

Ort: Raum Nr. C210 UZA4

Anmeldung (und evt. Ummeldung): läuft vom 3.11.-9.11.

unter: [www.unet.univie.ac.at/~a0307997/](http://www.unet.univie.ac.at/~a0307997/)

Log in: mit Ihrer Unet E-Mail Adresse + Matrikelnummer

Da für die Studie jede einzelne Person wichtig ist, würde es mich freuen, wenn Sie sich doch noch zu einer Teilnahme entschließen würden. Ihre Testergebnisse werden selbstverständlich vertraulich behandelt und nicht an Dritte weitergegeben

Mit freundlichen Grüßen

Elisabeth Unterhofer

P.S. Sollten Sie Probleme bei der Anmeldung haben, bitte geben Sie mir Bescheid. Bei erfolgreicher Anmeldung, erhalten Sie eine automatische Bestätigung via E-Mail.

Betreff: Bitte um Teilnahme an Testung zum logischen Denken

Liebe/r Frau/Herr XY

Vor einigen Tagen haben Sie eine E-Mail erhalten, mit dem Hinweis, dass an der Fakultät für Mathematik, Testungen zum logischen Denken im Rahmen einer Diplomarbeit aus Psychologie durchgeführt werden und dafür Mathematikstudenten gesucht werden. Da Sie heute nicht zur Testung erschienen sind, möchte ich Sie erneut um eine Teilnahme bitten:

Zeit: 10.-12. November, 11-17 Uhr- genaue Termine siehe Internetseite unten

Dauer: max.75 min

Ort: Raum Nr. C210 UZA4

Anmeldung (und evt. Ummeldung): läuft vom 3.11.-9.11.

unter: [www.unet.univie.ac.at/~a0307997/](http://www.unet.univie.ac.at/~a0307997/)

Log in: mit Ihrer Unet E-Mail Adresse + Matrikelnummer

Da für die Studie jede einzelne Person wichtig ist, würde es mich freuen, wenn Sie sich doch noch zu einer Teilnahme entschließen würden. Ihre Testergebnisse werden selbstverständlich vertraulich behandelt und nicht an Dritte weitergegeben

Mit freundlichen Grüßen

Elisabeth Unterhofer

P.S. Sollten Sie Probleme bei der Anmeldung haben, bitte geben Sie mir Bescheid. Bei erfolgreicher Anmeldung, erhalten Sie eine automatische Bestätigung via E-Mail.

## Persönlicher Aufruf

Hallo, mein Name ist Elisabeth Unterhofer. Ich habe an Sie alle in den letzten Wochen E-Mails verschickt und Sie gebeten an einer Testung teilzunehmen, die ich im Rahmen meiner Diplomarbeit aus Psychologie, an der Fakultät für Mathematik durchführe.

Nochmal ganz kurz: in der Studie geht es um einen Test, der logisches Denken erfasst und ich suche dafür dringend Mathematikstudenten, die daran teilnehmen. In diesem Seminar haben von den  $x$  Studenten  $y$  noch nicht teilgenommen; nächste Woche finden noch einmal Testungen statt und ich würde diese  $x$  Studenten daher bitten, doch noch an einer Testung teilzunehmen, da es wirklich wichtig wäre\*.

Die Testungen finden wieder von Montag- Mittwoch von 11.00-18.15 und dauern max 75 min.

Die Anmeldung erfolgt nicht mehr online, sondern direkt mit diesem Zeitplan auf dem Sie sich eintragen können. Zu jedem Zeitpunkt können sich bis zu 13 Personen anmelden, bei jedem Termin hängt ein Zettel, den sie runterreißen können, auf dem steht das Testdatum und die Uhrzeit drauf, damit Sie es nicht vergessen. .

Die Testung findet dann im Raum Nr. C210A, UZA 4 statt.

**LV- Leiter:** „Frau Unterhofer wird in ihrer Diplomarbeit auch von Seiten der Fakultät unterstützt, daher hat sie auch einen Raum für die Testungen erhalten und die Möglichkeit bekommen mit Ihnen Kontakt aufzunehmen. Ich bitte deshalb die Personen, die noch nicht an der Testung teilgenommen haben, dies zu tun und sich anzumelden.“

---

\* Falls niemand teilgenommen hat: In diesem Seminar hat leider niemand teilgenommen. Nächste Woche finden noch einmal Testungen statt, daher würde ich Sie alle bitten doch noch an einer Testung teilzunehmen, da es wirklich wichtig wäre.

## **Instruktion Testung**

Ich möchte alle zur Testung willkommen heißen und mich für die Teilnahme bedanken.

Bevor wir anfangen hätte ich eine Bitte an Sie: Geben Sie keine Informationen über die Testung, an andere Studenten weiter. Im schlimmsten Fall, könnten sonst die Testergebnisse der nachfolgenden Personen dadurch beeinflusst werden und es kommt zu einer Verzerrung in den Daten. Da ich das unbedingt vermeiden möchte, bitte ich Sie nochmals sich daran zu halten und keine Informationen weiterzugeben. Danke

Wir werden jetzt einzeln zu Ihnen kommen und den Test starten. Der Test besteht aus 3 verschiedenen Aufgaben.

Falls Sie eine Instruktion nicht verstehen sollten oder Fragen haben, dann melden Sie sich einfach.

Ich versichere Ihnen nochmals, dass die Testergebnisse vertraulich behandelt werden und nicht an Dritte weitergegeben werden.

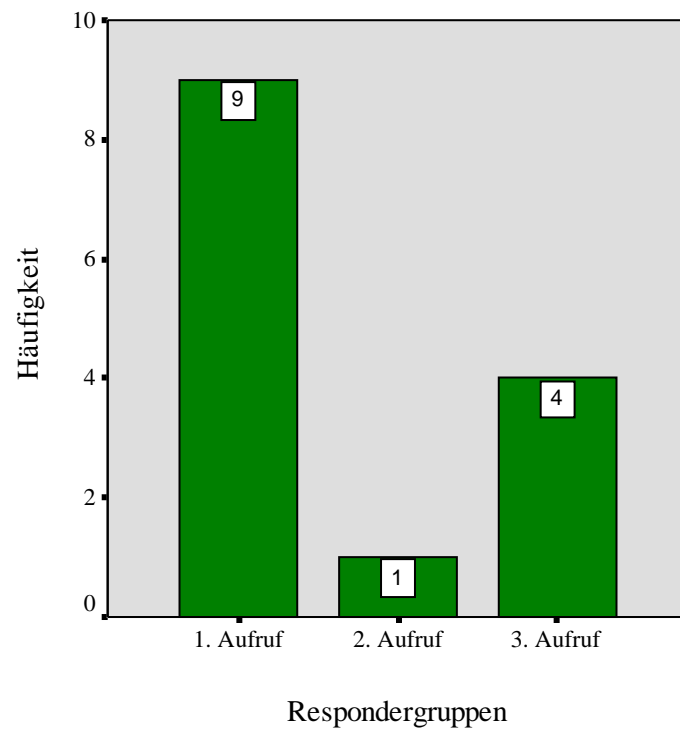
Nach Beendigung der Testung:

„Vielen Dank noch mal für die Teilnahme. Falls Sie Interesse an einer Rückmeldung über Ihre Testergebnisse haben, können Sie sich nach Beendigung der Studie ab Anfang März, bei mir via E-Mail melden.“

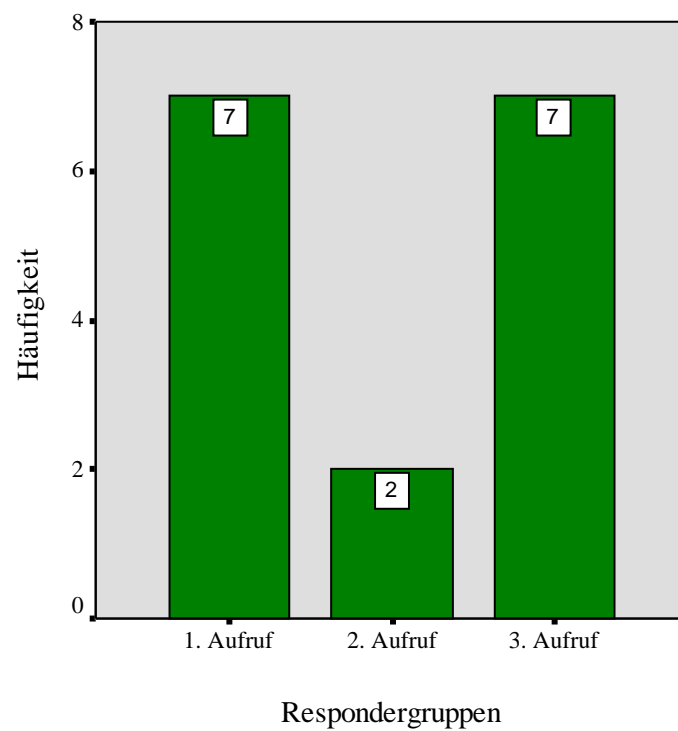
### **13.2. Anhang B - Tabellen und Abbildungen**

- Beschreibung der Stichprobe II - Männer - Variante 1
- Beschreibung der Stichprobe II - Frauen - Variante 1
- Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1a
- Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1b
- Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1c
- Beschreibung Stichprobe II - Männer - Variante 2
- Beschreibung Stichprobe II - Frauen - Variante 2

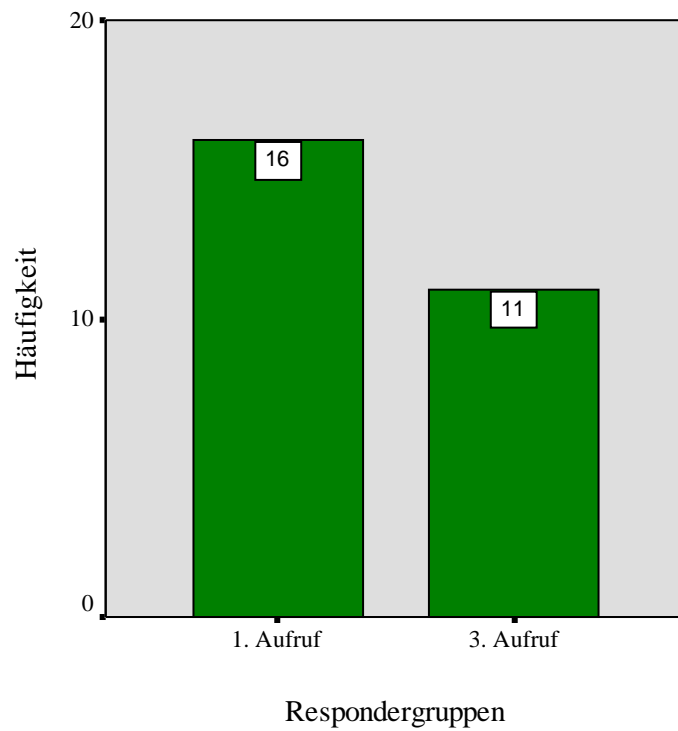
### Beschreibung der Stichprobe II - Männer - Variante 1



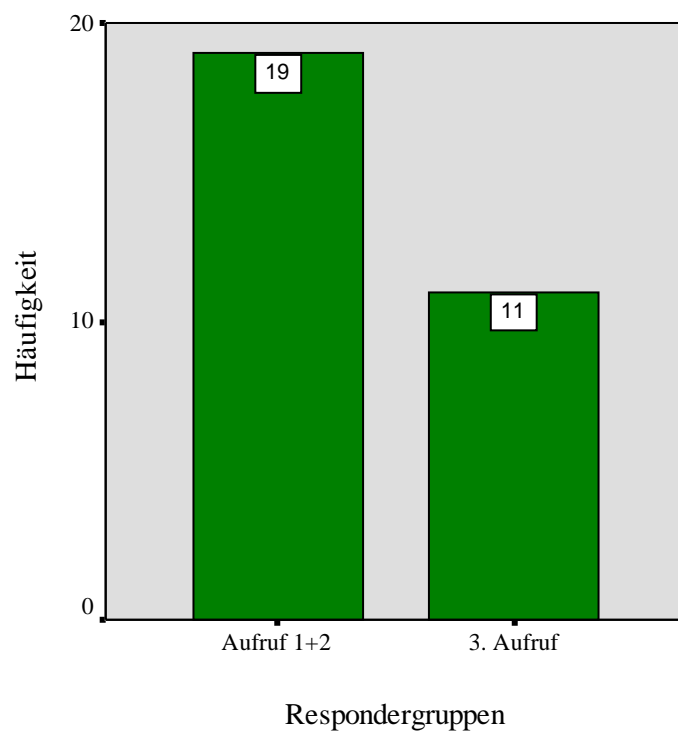
### Beschreibung der Stichprobe II - Frauen - Variante 1



### Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1a

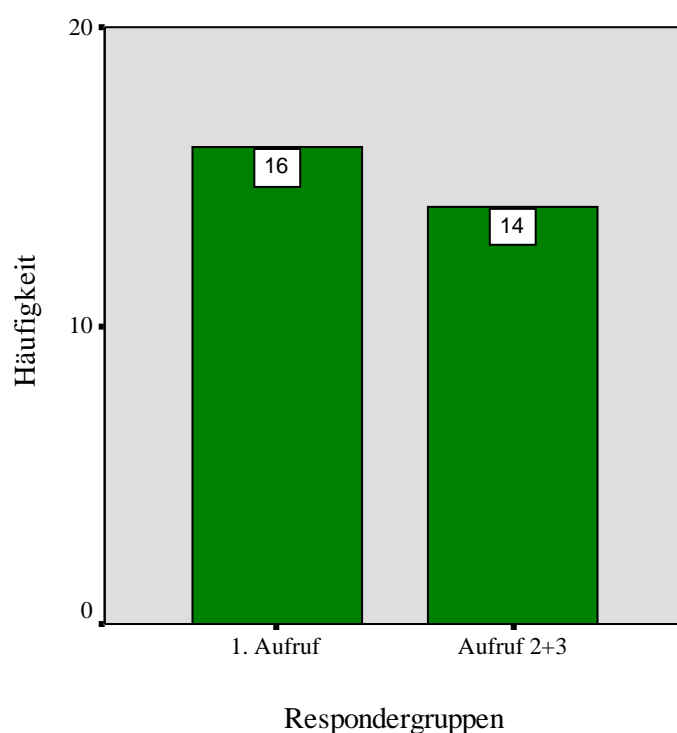


### Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1b





## Beschreibung Stichprobe II - Gesamtstichprobe Aufrufe - Variante 1c



## Beschreibung Stichprobe II - Männer - Variante 2

	Häufigkeit	Prozent
<b>Umkod. Pers.</b>	7	50,0
<b>1. Aufruf</b>	2	14,3
<b>2. Aufruf</b>	1	7,1
<b>3. Aufruf</b>	4	28,6
<b>Gesamt</b>	14	100,0

## Beschreibung Stichprobe II - Frauen - Variante 2

	Häufigkeit	Prozent
<b>Umkod. Pers.</b>	6	37,5
<b>1. Aufruf</b>	1	6,2
<b>2. Aufruf</b>	2	12,5
<b>3. Aufruf</b>	7	43,8
<b>Gesamt</b>	16	100,0

# LEBENS LAUF

## Persönliche Daten

Name	Elisabeth Unterhofer
E-Mail-Adresse:	a0307997@unet.univie.ac.at
Familienstand	ledig
Staatsangehörigkeit	italienisch
Geburtsdaten:	28. März 1984, Brixen/ Südtirol

## Schulische Ausbildung

1998-2003	Pädagogisches Gymnasium Bruneck, Abschluss: Matura
1995-1998	Mittelschule Sand in Taufers
1990-1995	Grundschule Sand in Taufers

## Studium

Derzeit	Diplomarbeit im Bereich Psychologische Diagnostik
Februar 06	Abschluss des ersten Abschnittes
WS 2003	Beginn des Studiums der Psychologie an der Universität Wien

## Zusätzliche Weiterbildungen und praktische Erfahrungen im Bereich der Psychologie

April 2009	Studienassistentin am Arbeitsbereich für Psychologische Diagnostik
November 2008	Abschluss des Zertifizierungskurses für das AID 2
Mai 2008	Teilnahme an der „Einführung in das Softwarepaket Extended Rasch modeling (eRm)“ zu 5 h
WS 2007/08	Im Rahmen des „Praktikums zum psychologischen Diagnostizieren“: Durchführung einer Eignungsdiagnostik für das Psychologiestudium bei 2 Studenten. Diese umfasste ein Erstgespräch, die Durchführung der Testung sowie ein abschließendes Beratungsgespräch.
Herbst 2007	6 wöchiges Praktikum an der Ambulanz der Psychiatrie des Kindes- und Jugendalters, AKH Wien.

Schwerpunkt: Anwendung und Auswertung  
psychologisch- diagnostischer Verfahren:  
Persönlichkeits- und Leistungsdiagnostik

2006/7

Teilnahme an einer Gruppenselbsterfahrung im  
Ausmaß von 50 h, bei Mag. Annelies Weber, Autogene  
Psychotherapeutin

1999 & 2001

Teilnahme an zweiwöchigem Englischsprachkurs in  
Irland

### **Weitere Kenntnisse**

Sprachkenntnisse

Deutsch (Muttersprache)  
Italienisch  
Englisch  
Latein (kleines Latinum)

EDV-Kenntnisse

Kurs zum Zehn-Finger-System, Textverarbeitung  
(Word), Präsentationsprogramme (PowerPoint),  
Internet und E-Mailing, Statistikauswertung (SPSS),  
CorelDraw, Excel Grundkenntnisse, Extended Rasch  
modeling (eRm)